OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: July 16, 2003, 22:42:04 ; Search time 7008.05 Seconds
(without alignments)
17408.415 Million cell updates/sec

Title: US-09-720-086-1
Perfect score: 4192
Sequence: 1 gaattccggcctgctgccgg..........taaaaaacatacaaaatgt 4192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455142878 residues

Total number of hits satisfying chosen parameters: 4109280

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :
```
GenEmbl:*
 1: gb_ba:*
 2: gb_htg:*
 3: gb_in:*
 4: gb_om:*
 5: gb_ov:*
 6: gb_pat:*
 7: gb_ph:*
 8: gb_pl:*
 9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4191	100.0	4192	10	AF068625	AF068625 Mus muscu
	2	3982.8	95.0	4094	10	BC007466	BC007466 Mus muscu
	3	2877.4	68.6	4258	9	AF331856	AF331856 Homo sapi
	4	2394.6	57.1	3005	9	AF067972	AF067972 Homo sapi
	5	1551.6	37.0	2191	9	AK025230	AK025230 Homo sapl
	6	1369.4	32.7	204534	2	AC118195	AC118195 Mus muscu
	7	1369.4	32.7	208874	2	AC111092	AC111092 Mus muscu
	8	1331.8	31.8	123336	2	AC120824	AC120824 Rattus no
c	9	1331.8	31.8	176697	2	AC112586	AC112586 Rattus no
c	10	1171.2	27.9	1758	9	BC018214	BC018214 Homo sapi
c	11	1164.8	27.8	117264	2	AC103454	AC103454 Rattus no
c	12	1055.2	25.2	168651	2	AC009474	AC009474 Homo sapi
c	13	1015.8	24.2	119630	2	AC112040	AC112040 Rattus no
c	14	962	22.9	4135	10	AF068627	AF068627 Mus muscu
	15	796.2	19.0	4163	10	AF151974	AF151974 Mus muscu
	16	794.6	19.0	4278	10	AF151970	AF151970 Mus muscu
	17	794.6	19.0	4267	9	AF176228	AF176228 Homo sapi
	18	789.6	18.8	4195	10	AF068626	AF068626 Mus muscu
	19	759.4	18.1	4223	10	AF151973	AF151973 Mus muscu
	20	757.8	18.1	4338	10	AF151969	AF151969 Mus muscu
	21	757.8	18.1	4145	9	AF156488	AF156488 Homo sapi
c	22	749.8	17.9	4335	9	AF331857	AF331857 Homo sapi
c	23	749.8	17.9	197652	9	AC012074	AC012074 Homo sapi
c	24	640.4	15.3	210269	2	AC116459	AC116459 Mus muscu
c	25	616.4	14.7	3946	10	AF068628	AF068628 Mus muscu
c	26	604.8	14.4	3974	10	AF151976	AF151976 Mus muscu
	27	603.2	14.4	4089	10	AF151971	AF151971 Mus muscu
	28	603.2	14.4	3897	9	AF156487	AF156487 Homo sapi
	29	597.6	14.3	4006	10	AY078427	AY078427 Mus muscu
	30	568	13.5	4034	10	AF151975	AF151975 Mus muscu
	31	566.4	13.5	4149	10	AF151972	AF151972 Mus muscu
	32	566.4	13.5	119630	2	AF135438	AF135438 Danio rer
	33	564.6	13.5	2848	5	AK001191	AK001191 Homo sapi
	34	531.8	12.7	3017	9	AF129267	AF129267 Homo sapi
	35	482.8	11.5	2127	9	AC111734	AC111734 Rattus no
	36	417.6	10.0	167568	2	AR129189	AR129189 Sequence
	37	416.4	9.9	2077	6	AF129268	AF129268 Homo sapi
	38	324.6	7.7	2057	9	AC118195	AC118195 Mus muscu
c	39	302.8	7.2	204534	2	AC111092	AC111092 Mus muscu
c	40	273	6.5	208874	2	AC119009	AC119009 Rattus no
c	41	273	6.5	185413	9	AF129269	AF129269 Homo sapi
c	42	258	6.2	2008	2	AC101453	AC101453 Mus muscu
	43	255.2	6.1	64045	9	AF194032	AF194032 Homo sapi
	44	216.4	5.2	1705	9	BC002560	BC002560 Homo sapi
	45	141	3.4	1397	9		
		140	3.3				

ALIGNMENTS

```
RESULT 1
AF068625
AF068625
LOCUS       AF068625                4192 bp    mRNA    linear   ROD 06-DEC-1999
DEFINITION  Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,
            complete cds.
ACCESSION   AF068625
VERSION     AF068625.2  GI:6449467
KEYWORDS    .
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4192)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE     Cloning and characterization of a family of novel mammalian DNA
```

```
JOURNAL       (cytosine-5) methyltransferases
MEDLINE       Nat. Genet. 19 (3), 219-220 (1998)
PUBMED        98324766
REFERENCE     9662389
  AUTHORS     2 (bases 1 to 4192)
  TITLE       Xie,S., Okano,M. and Li,E.
  JOURNAL     Direct Submission
              Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
              Charlestown, MA 02129, USA
REFERENCE     3 (bases 1 to 4192)
  AUTHORS     Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE       Direct Submission
  JOURNAL     Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th street,
              Charlestown, MA 02129, USA
  REMARK      Sequence update by submitter
COMMENT       On Nov 18, 1999 this sequence version replaced gi:3327977.
FEATURES             Location/Qualifiers
     source          1..4192
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /map="4.0 cM"
     gene            1..4192
                     /gene="Dnmt3a"
     misc_feature    1..4192
                     /gene="Dnmt3a"
                     /note="similar to EST sequences deposited in GenBank
                     Accession Numbers AA052791, AA111043, AA154890, AA240794,
                     AA756553, W58998, W59299, W91664, and W91665"
     CDS             217..2943
                     /gene="Dnmt3a"
                     /function="de novo DNA methylation"
                     /note="contains Cys-rich region"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3A"
                     /protein_id="AAC40177.2"
                     /db_xref="GI:6449468"
                     /translation="MPSSGPGDTSSSSLEREDDRKEGEEQENRGKEEROEPSATARK
                     VGRPGRKRKHPPVESSSDTPKDPAVTTKSOPMAQDSGPSDLLPNGDLEKRSEPQEEGS
                     PAAGQKGAPAEGEGTETPEEASRAVENGCCVTKEGRGASAGEGKEQKQTNIESMKME
                     GSRGRLRGLGWESSLRQRPMPRLTFQAGDPYIISKRKRDEWLARWKREAERKAKVIA
                     VMNAVEENQASGESQKVEEASPAVOQPTDPASPTVATTPEPVGGDAGDKNATKAADD
                     EPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDGKFSV
                     VCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFPACHDSDESDSG
                     KAVEVQNKQMIEWALGGFQPSGPRKLEPPEEEKNPYKEVYTDMNVEPEANAVAPPPPA
                     KPRKSTTEKPKVKEIIDERTRERLVYEVRQKCRNIEDICICSGSLNVTLEHPLFIGG
                     MCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFVECVDLLVGPG
                     AAQAAIKEDPWNCYMCGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYPPVPA
                     EKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGD
                     VRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEYRLLHDARPK
                     EGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMN
                     RPLASTVNDKLELQECLEHGRIAKFSKVRTIITRSNSIKQCKDQHFPVFMNEKEDILW
                     CTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACY"
ORIGIN
BASE COUNT   1094 a   1072 c   1186 g    840 t
```

```
Query Match        100.0%;    Score 4191;    DB 10;    Length 4192;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 4191; Conservative  0;  Mismatches 0;  Indels  0;  Gaps  0;

Qy          1 GAATTCCGGCCTGCTGCCGCGGCCGGCCGCCGCGGGCCGCGCGGCCACACGGCAGAGCCGCCTGAA   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCCGGCCTGCTGCCGCGGCCGGCCGCCGCGGGCCGCGCGGCCACACGGCAGAGCCGCCTGAA   60

Qy         61 GCCCAGCGGCTGAGGCTGCACTTTTCCGAGGGCTTGACATCAGGGTCTCTATGTTTAAGTCTT  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCCAGCGGCTGAGGCTGCACTTTTCCGAGGGCTTGACATCAGGGTCTCTATGTTTAAGTCTT  120

Qy        121 AGCTCTTGCTTACAAAGACCACGGCAAATTCCTTCTTCTCTGGAAGCCCTCGCAGCCCCCACACGG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGCTCTTGCTTACAAAGACCACGGCAAATTCCTTCTTCTCTGGAAGCCCTCGCAGCCCCCACACGG  180

Qy        181 CCCTCGCAGCCCCACGCCTACTGCCCAGCAAATGCCCTCCAGCGGCGGCCCCGGGGAC  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCTCGCAGCCCCACGCCTACTGCCCAGCAAATGCCCTCCAGCGGCGGCCCCGGGGAC  240

Db        241 ACCAGCAGCTCCTCTCTGGAGCGCGGAGGAGTGATCGCAGCAATGCCCTCACGCGCCCCGGGGAC  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        241 ACCAGCAGCTCCTCTCTGGAGCGCGGAGGAGTGATCGCAGCAATGCCCTCACGCGCCCCGGGGAC  300

Qy        301 AACCGTGGCAAGGAAGAGCGCGAGCAGGAGCCCAGCGCCACGGCGCCCAGCGGAGGTGGGAGGCCT  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACCGTGGCAAGGAAGAGCGCGAGCAGGAGCCCAGCGCCACGGCGCCCAGCGGAGGTGGGAGGCCT  360

Db        361 GGCCGGAAGCGCAAGCACCCCACCGGTTGGAAAGCAGCAGTGACACCCCAAGGACCCAGCAGTC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        361 GGCCGGAAGCGCAAGCACCCCACCGGTTGGAAAGCAGCAGTGACACCCCAAGGACCCAGCAGTC  420

Qy        421 ACCACCAAGTCTCAGCCCATGGCCCAGGACTCTGCCCCTCAGATCTGCTTACCCAATGGA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACCACCAAGTCTCAGCCCATGGCCCAGGACTCTGCCCCTCAGATCTGCTTACCCAATGGA  480

Qy        481 GACTTGGAGAAACGGAGCGAGTGAACCCCAACCTGAAGGAGGGGAGCCCAGCCTGCAGCGGCAGAAG  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTTGGAGAAACGGAGCGAGTGAACCCCAACCTGAAGGAGGGGAGCCCAGCCTGCAGCGGCAGAAG  540

Qy        541 GGTGGGGCGCCCCACCTGAAGGAGGAGGAAACTGAGACCCCAACCAGAAGAGCCTTCCAGACCTGTG  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGTGGGGCGCCCCACCTGAAGGAGGAGGAAACTGAGACCCCAACCAGAAGAGCCTTCCAGACCTGTG  600

Db        601 GAGAATGGCTGCTGTGTGACCAAGGAAGGGCGTGGAGGAGCCCTCGTGCAGGAGAGGGCAAAGAA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        601 GAGAATGGCTGCTGTGTGACCAAGGAAGGGCGTGGAGGAGCCCTCGTGCAGGAGAGGGCAAAGAA  660

Qy        661 CAGAAGCAGACCAACATCGAATCCATGAAAATGGAGGGGCTCCCGGCGGCCCGACTGCGAGGT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAGAAGCAGACCAACATCGAATCCATGAAAATGGAGGGGCTCCCGGCGGCCCGACTGCGAGGT  720

Db        721 GGCTTGGGCTGGGAGTCCAGCCTCCGTCAGCGGACCATGCCAAGACTCACCTTCCAGGCA  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        721 GGCTTGGGCTGGGAGTCCAGCCTCCGTCAGCGGACCATGCCAAGACTCACCTTCCAGGCA  780

Qy        781 GGGGACCCCTACTACATCAGCAAAAGGAAAAGGGATGAGTGGCTGGCACGTTGGAAAAGG  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGGGACCCCTACTACATCAGCAAAAGGAAAAGGGATGAGTGGCTGGCACGTTGGAAAAGG  840

Db        841 GAGGCTGAGAAGAAAGCCAAAGTAATTGCAGTAATGAATGCTGTGGAAGAGAACCAGGCC  900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        841 GAGGCTGAGAAGAAAGCCAAAGTAATTGCAGTAATGAATGCTGTGGAAGAGAACCAGGCC  900

Qy        901 TCTGGAGAGTCTCAGAAAGGTGGAGGAGGCCAGCCCTCCTGCTGTGCAGCACCCACGGAC  960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCTGGAGAGTCTCAGAAAGGTGGAGGAGGCCAGCCCTCCTGCTGTGCAGCACCCACGGAC  960

Qy        961 CCTGCTTCTCCGACTGTGGCCACCACCCCTCAGCCAGTAGGAGGGGATGCTGGGGACAAG  1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCTGCTTCTCCGACTGTGGCCACCACCCCTCAGCCAGTAGGAGGGGATGCTGGGGACAAG  1020

Db       1021 AATGCTACCAAAGCAGCCGACGATGAGCCTGAGTATGAGGATGGCCGGGGCTTTGGCATT  1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1021 AATGCTACCAAAGCAGCCGACGATGAGCCTGAGTATGAGGATGGCCGGGGCTTTGGCATT  1080

Qy       1081 GGAGAGCTGGTGTGGGGGAAACTTCGGGGCTTCTCTCGGTGGCCAGGCCGAATTGTGTCT  1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGAGAGCTGGTGTGGGGGAAACTTCGGGGCTTCTCTCGGTGGCCAGGCCGAATTGTGTCT  1140

Qy       1141 TGGTGGATGACAGGCCGGAGCCGAGCACTCGGCTCGGCTCGGGTCATGTGGTTCGGA  1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGGTGGATGACAGGCCGGAGCCGAGCACTCGGCTCGGCTCGGGTCATGTGGTTCGGA  1200

Db       1201 GATGGCAAGTTCTCAGTGGTGTGTGTGGAGAAGCTCATGCCGCTCAGCTCCTTCTGCACT  1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1201 GATGGCAAGTTCTCAGTGGTGTGTGTGGAGAAGCTCATGCCGCTCAGCTCCTTCTGCACT  1260

Qy       1261 GCATTCCACCAGGCCACCTACAACAAGCAGCCCATGTACCGCAAAGCCATCTACGAAGTC  1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1261 GCATTCCACCAGGCCACCTACAACAAGCAGCCCATGTACCGCAAAGCCATCTACGAAGTC 1320

Qy   1321 CTCCAGGTGGCCAGCGAGCCGTGTCCCAGGGAAGCTGTTTCCCAGCTTGCCAGCACGTGATGAA 1380
          ||||||||||| |||||| |||| ||| |||||||||||||||||||||||  |||  |||||||
Db   1321 CTCCAGGTGGGCCAGCAGCCAGCCCGTGCCCGGGAAGCCTGTTTTCCCAGTCGATGACAGTGATGAA 1380

Qy   1381 AGTGACAGTGGGCAAGGCTGTGGAAGTGCAGAAACAAGCAGATGATTGAATGGGCCCTCGGT 1440
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1381 AGTGACAGTGGGCAAGGCTGTGGAAGTGCAGAACAAGCAGATGATTGAATGGGCCCTCGGT 1440

Qy   1441 GGCTTCCAGCCCTGGTCCTTAGGGGCCTCTGGAGCCGCCACCGAGAAGAAGAGAATCCTTAC 1500
          |||||||||||| ||| ||| |||| ||  ||||||| ||| ||||||   |||| |||||
Db   1441 GGCTTCCAGCCCTCGGGTCCTAAGGGCCTGGAGCCGACCAAAGAGAAATCCTTAC 1500

Qy   1501 AAGGAAGTTTACACCGACATGTGGGGTGGGAGCCTGAAGCAGCTGCTTACGCCCCACCCCCCA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AAGGAAGTTTACACCGACATGTGGGGTGGGAGCCTGAAGCAGCTGCTTACGCCCCACCCCCCA 1560

Qy   1561 CCAGCCAAGAAAACCCAGANAGAGACCACACAGAGAAACCTAAGGTCAAGGAGGAATCATTGAT 1620
          ||||||||||||||||||| ||||  ||||||||||||||||||||||||||||||||||||
Db   1561 CCAGCCAAGAAAACCCAGANAGAGCACACAGAGAAACCTAAGGTCAAGGAGGAATCATTGAT 1620

Qy   1621 GAGCGCACAAGGGGAGCGGGCCTGGTGTTGATGAGGTGCGCCAGAAGTGCAGAAACATCGAGGAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GAGCGCACAAGGGGAGCGGGCCTGGTGTTGATGAGGTGCGCCAGAAGTGCAGAAACATCGAGGAC 1680

Qy   1681 ATTTGTTGTATCTCATGTGGGAGCCCTCAAATGTCACCCACCCACTTCCATTGGAGGC 1740
          |||||| ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1681 ATTTGTATCTCATGTGGGAGCCCTCAAATGTCACCCACCCACTTCATTGGAGGC 1740

Qy   1741 ATGTGCCAGAAACTGTAAGAAACTGCTTCTTCTTGGATGTGCTTACCAGTATGACCAGATGTGGG 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ATGTGCCAGAAACTGTAAGAAACTGCTTCTTCTTGGATGTGCTTACCAGTATGACCAGATGTGGG 1800

Qy   1801 TACCAGTCCTATTGCACCATCTGCTGTGTGGGAGTGCTCATCGTGTGGGAACAAC 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TACCAGTCCTATTGCACCATCTGCTGTGTGGGAGTGCTCATCGTGTGGGAACAAC 1860

Qy   1861 AACTGCTGCAGGTGCTTTGTGTGGGATCTCTTCGGTGGGGCCAGGAGCTGCT 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AACTGCTGCAGGTGCTTTGTGTGGGATCTCTTCGGTGGGGCCAGGAGCTGCT 1920

Qy   1921 CAGGCAGCCCATTAAGGAAGAGACCCCCTGGGAACTGCTACATGTGGGCGCCATGAGGGCCACCTAT 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CAGGCAGCCCATTAAGGAAGAGACCCCCTGGGAACTGCTACATGTGGGCGCCATGAGGGCCACCTAT 1980

Qy   1981 GGGCTGCTTGCGGAAGGACGGGGAAGACTGGCCTTCTCGACTCCAGATGTCTGTTTTGCCAATAAC 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GGGCTGCTTGCGGAAGGACGGGGAAGACTGGCCTTCTCGACTCCAGATGTCTGTTTTGCCAATAAC 2040

Qy   2041 CATGACCAGGGAAATTTGACCCCCAAAAGGTTTACCCAGTTCTTTGCCAATAAC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CATGACCAGGGAAATTTGACCCCCAAAAGGTTTACCCAGTTCTTTGCCAATAAC 2100

Qy   2101 CCCATCCGCGTGCTGTCTCTCTTTCGATGGGATTGGGCTACCAGGGGCTCCTGTGGTGCTGAAGGAC 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CCCATCCGCGTGCTGTCTCTCTTTCGATGGGATTGGGCTACCAGGGGCTCCTGTGGTGCTGAAGGAC 2160

Qy   2161 CTGGGCATCCAAGTGGACGCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTG 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CTGGGCATCCAAGTGGACGCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTG 2220

Qy   2221 GGCATGTGTGCGGCACCCAGGAGGACGTCCCGCAGCCGTCACACACAG 2280
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GGCATGTGTGCGGCACCCAGGAGGACGTCCCGCAGCCGTCACACACAG 2280

Qy   2281 AAGCATATCCAGGAGTGGGGGCCCATTCCGACCTGGTGTCCCCTGCAATGAC 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AAGCATATCCAGGAGTGGGGGCCCATTCCGACCTGGTGTCCCCTGCAATGAC 2340

Qy   2341 CTCTCCATTGTCAACCCTCCCGCAAGGGACTTTTATGAGGGTTAC 2400
          |||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTCTCCATTGTCAACCCTCCCGCAAGGGACTTTTATGAGGGTTAC 2400
```

```
Qy   2401 GAGTTCTTACCGCGCGCTGCCTCCATGATGCCGCGGCCGCCAAGGAGGAGGAGGAGGATGATCGCCCCGGATTTCTC 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GAGTTCTTACCGCGCGCTGCCTCCATGATGCCGCGGCCGCCAAGGAGGAGGAGGAGGATGATCGCCCCGGATTTCTC 2460

Qy   2461 TGGCTCTTTGAGAATGTTGGTGTGGCCATTGGGGTTAGTGACAAGAGGGGACATCTCGGCGATTT 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TGGCTCTTTGAGAATGTTGGTGTGGCCATTGGGGTTAGTGACAAGAGGGGACATCTCGGCGATTT 2520

Qy   2521 CTTGAGTGTAACCCCGTGATGATTGACGCCAAAGAGAAGTGTCTGCTGCACACAGGGCCGGT 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CTTGAGTCTAACCCCGTGATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGGCCCGT 2580

Qy   2581 TACTTCTGGGGTAACCTTCCTGGCATGACAGGCCCTTTGGCATCCACTGTGAATGATAAG 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TACTTCTGGGGTAACCTTCCTGGCATGACAGGCCATCCACTGTGAATGATAAG 2640

Qy   2641 CTGGAGCTGCAAGAGTGTCTGGACCACGGCCAGAAATAGCCCAAGTTCAGCAAGTGAGGACC 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 CTGGAGCTGCAAGAGTGTCTGGAGCACGGCCAGAAATAGCCAAGTTCAGGAGGACC 2700

Qy   2701 AACGAGAAGGAGGACAGCATCCTGTGGTGCACTGAAAATGGAAAAGGGTGTTGTTGGCTTCATG 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 AACGAGAAGGAGGACAGCATCCTGTGGTGCACTGAAAATGGAAAAGGGTGTTGTTGGCTTCATG 2760

Qy   2761 ATTACCCACCAGGTCAAACTCTATAAAGCAGGGCAAAGCACCAGCCATTTCCCCGTCTC 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 ATTACCCACCAGGTCAAACTCTATAAAGCAGGGCAAAGCACCAGCCATTTCCCCGTCTC 2820

Qy   2821 CACTACACGAGCGTCTCCAAACATGAGCCGCTGCTGCGGCGGCGCTGGGCGCCGATCG 2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CACTACACAGAGCGTCTCCAAACATGAGCCGCTGGCGGCGCTGGGCGCCGATCG 2880

Qy   2881 TGGAGCGTGCCGGTCATCCGGCCACCTCTTCCGCTCCCCTGAAGGAAATATTTGCTTGGTGTG 2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TGGAGCGTGCCGGTCATCCGGCCACCTCTTCCGCTCCCCTGAAGGAAATATTTGCTTGGTGTG 2940

Qy   2941 TAAGGGGACATGGGGGCAAAACTGAAGTAGTGATGAAGTAAAAACACAAACACAAC 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 TAAGGGGACATGGGGGCAAAACTGAAGTAGTGATGAAGTAAAAACACAAACACAAC 3000

Qy   3001 AAAAAACAAAAACAAAACAATAAAACAACAAGAAGACGAGAAAGTTCAGCACC 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 AAAAAACAAAAACAAAACAATAAAACAACAAGAAGACGAGAAAGTTCAGCACC 3060

Qy   3061 CAGAAGAGGAGAAAAAGGGTTGGACAATCACATCCTCGAGTTTTTCAGTGCTTATCTA 3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 CAGAAGACAAAAAGGAATTAAAGCAAACACACAGAGGGAAAACGCCGGAGGGCTTGGC 3120

Qy   3121 CTTGCAAAAGGGTTGGACATCACATCCTCGAGTTTTTCAGTGCTTATCTA 3180
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 CTTGCAAAAGGGTTGGACATCACATCCTCGAGTTTTTCAGTTCCTATCTA 3180

Qy   3181 AAAAGCAAAATAGGCCCCTCCCCTTCCCCTTTCCCGGTTCCTAGGAGGGCGGAACTTTTTGTT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AAAAGCAAAATAGGCCCCTCCCTTCTTCCCCGGTTCCTAGGAGGGCGGAACTTTTTGTT 3240

Qy   3241 TTCTACTCTTTTCAGAGGGGTTTTGTGTTTGTTTGTTTTGTTTTGCTGTGACTG 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTCTACTCTTTTCAGAGGGGTTTTGTGTTTGTTTGTTTTGTTTTGCTGTGACTG 3300

Qy   3301 AAACAACAGAGAGTTATTGCAGCAAAATCAGTAACAACAAAGTAGAAAATGCCTTGGAGAG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   3301 AAACAACAGAGAGTTATTGCAGCAAAATCAGTAACAACAAAGTAGAAAATGCCTTGGAGAG 3360

Qy   3361 GAAAGGGAGAGAGGGAAAAATTCTATAAAAAAACTTAAAAATATATTTGTTTTTTTTTTTCCT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 GAAAGGGAGAGAGGGAAAAATTCTATAAAAAACTTAAAAATATTTGTTTTTTTTTTTTCCT 3420

Qy   3421 TTCTATATATCTCTTTGGTTGTTGTCTCTAGCCTGATCAGATAGGAGCACAAAACAGGAAGAG 3480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 TTTCTATATATCTCTTTGGTTGTTGTCTCTAGCCTGATCAGATAGGAGCACAAAACAGGAAGAG 3480
```

LOCUS   BC007466   4094 bp   mRNA   linear   ROD 07-AUG-2002
DEFINITION Mus musculus, DNA methyltransferase 3A, mRNA, clone MGC:5662
            IMAGE:3492853, mRNA, complete cds.
ACCESSION   BC007466
VERSION   BC007466.1 GI:13938620
KEYWORDS   MGC.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4094)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil1.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            NIH-MGC Project URL: http://mgc.ncl.nih.gov

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 8 Row: n Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6681208.
FEATURES   Location/Qualifiers
     source   1..4094
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="CZECH II"
                     /clone="MGC:5662 IMAGE:3492853"
                     /tissue_type="Mammary tumor metastatized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMTV."
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /note="vector: pCMV-SPORT6"
     CDS   245..2971
                     /codon_start=1
                     /product="DNA methyltransferase 3A"
                     /protein_id="AAH07466.1"
                     /db_xref="GI:13938621"
                     /db_xref="LocusID:13435"
                     /translation="MPSSGPGDTSSSSLEREDDRKEGEEQEENNGKEERQEPSATARK
                     VGRPGRKRKHPVESSDTPKDPAVTTKSQPMAQDSGPSDLLPNGDLEKRSEPQPEEGS
                     PAAGRKGGAPAEGEGTETPPEASRAVENGCCVTEKGRCASAGESKEQKPTNIESMKME
                     GSRGRLRGGLGWESSLRQRPMPRLLFQAGDPYYISKRKRDEWLARWKREAEKKAKVIA
                     VMNAVEENQASGESQKVEEASPPAVQQPTDPASPTVATPEPVGGDAGDKNATKAADD
                     EPEYEDGRDGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDGKFSV
                     VCVEKLMPLSSFCSARHQATYNKQPMYRKAIYEVLQVASSRAGKLFPACHDSDESDSG
                     KAVEVQNKQMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMKWVEPEAAAYAPPPA
                     KKPRKSTTEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVTLEHPLFIGG
                     MCQNCKNCFLECAYQYDDDGYSYCTICCGREVLMCGNNNCCRCFCVECVDLLVGPG
                     AAQAAIKEDPWNCYMCGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPRKVYPPVPA
                     EKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGD
                     VRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEYYRLLHDARPK
                     EGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVHDAKEVSAAHRARYFWGNLPGMN
                     RPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDILW
                     CTEMRVFGFPVHYTDVSNMSRLARORLLGRSWSVPVIRHLFAPLKEYFACV"
BASE COUNT   1067 a   1164 c   806 g   1164 g   806 t
ORIGIN

RESULT 2
BC007466
LOCUS
DEFINITION Mus musculus, DNA methyltransferase 3A, mRNA, clone MGC:5662
            IMAGE:3492853, mRNA, complete cds.

Query Match   95.0%; Score 3982.8; DB 10; Length 4094;
Best Local Similarity   99.6%; Pred. No. 0;
Matches 4035; Conservative   0; Mismatches   12; Indels   5; Gaps   4;

Qy   3481 AATAGAGAACCCTCGGGAGGCAGAGTCTCCTCTCCCACCCCCGGAGCAGTCTCAACAGCACC 3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 AATAGAGAACCCTCGGGAGGCAGAGTCTCCTCTCCCACCCCCGGAGCAGTCTCAACAGCACC 3540

Qy   3541 ATTCCTGGTCATGCAAAACAGAACCCAAACTAGCACCAGCAGGGAACACCACACC 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 ATTCCTGGTCATGCAAAACAGAACCCAAACTAGCACCAGCAGGGAACACCACACC 3600

Qy   3601 AGACACTTTCTACAGTATTTCAGGTGCCTACCACAGGAAAAACCTTGAAGAAAAACCAGTT 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 AGACACTTTCTACAGTATTTCAGGTGCCTACCACAGGAAAAACCTTGAAGAAAAACCAGTT 3660

Qy   3661 TCTAGAAGCCCCGCTGTTGTTACCTCTTGTTTACAGTTATATATATATATGATAGATAT 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TCTAGAAGCCCCGCTGTTGTTACCTCTTGTTTACAGTTTATATATATATATGATAGATAT 3720

Qy   3721 ATATATATTAAAAGGTACTGTTAACTACTGTACATCCGACTTCATAAATGCTTTCAAA 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 ATATATATTAAAAGGTACTGTTAACTACTGTACATCCGACTTCATAAATGGTGCTTTCAAA 3780

Qy   3781 ACAGCGAGATGAAGCAAAGAACACATCAGCTTCCGCCCTTCGTGCAAAGGGTTTCAGCC 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 ACAGCGAGATGAAGCAAAGAACATCAGCTTCCCCCTGTGCAAAGGGTTTCAGCC 3840

Qy   3841 CAGGATGGGGACGAGGGGAGCAGCTGAGGGGGTTTTAACAAACTGAAGGATGACCCATAT 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 CAGGATGGGGACGAGGGGAGCAGCTGGAGGGGTTTTAACAAACTGAAGGATGACCCATAT 3900

Qy   3901 CACCCCCCCCCCCTGCCCCATGCCTAGCTTCACCTGCCAAAAAGGGCTCAGCTGAGGGTG 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 CACCCCCCCCCCTGCCCCATGCCTAGCTTCACCTGCCAAAAAGGGCTCAGCTGAGGGTG 3960

Qy   3961 GTCGGACCCTGGGGAAGCTGAGTGTGGAATTTATCCAGACTCGCGTGCAATAACCTTAGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GTCGGACCCTGGGGAAGCTGAGTGTGGAATTTATCCAGACTCGCGTGCAATAACCTTAGA 4020

Qy   4021 ATATGAATCTAAAATGACTGCCTCAGAAAAAATGGCTTGAGAAAAACATTGTCCCTGATTTT 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 ATATGAATCTAAAATGACTGCCTCAGAAAAAATGGCTTGAGAAAAACATTGTCCCTGATTTT 4080

Qy   4081 GAATTCGTTCAGCCACGTTGAAAGGCCCCTTGTGGGGATTCAGAAATATTCCAGAGTGAGGGAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAATTCGTTCAGCCACGTTGAAAGGCCCCTTGTGGGGATCAGAAATATTCCAGAGTGAGGGAA 4140

Qy   4141 AGTGACCCGCCCATTAACCCCCNCCTGGAGCAAATAAAAAAAAACATACAAAAAATGT 4192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 AGTGACCCGCCCATTAACCCCCNCCTGGAGCAAATAAAAAAAAACATACAAAAAATGT 4192

Qy   16 GCCGGGGCCCGCCCGCCGACCCGGCCGGGCCCACACGGCAGAGCCGCCTGAAGCCCACGCTGCAGGC 75
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16 GCCGGGGCCCGCCCGCCGACCCGGCCGGGCCCACACGGCAGAGCCGCCTGAAGCCCACGCTGCAGGC 75

Qy   44 GCCGGGGCCCGCCCGACCCCGCCCGGCCCACACGGCAGAGCCGCCTGAAGCCCACGCTGCAGGC 103
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44 GCCGGGGCCCGCCCGACCCCGCCCGGCCCACACGGCAGAGCCGCCTGAAGCCCACGCTGCAGGC 103

Qy   76 TGCCACTTTTCCGAGGGCTTGACATCAGGGTGTCTATGTTTAAGTCTTTAGCTCTTGCTTACAA 135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 TGCCACTTTTCCGAGGGCTTGACATCAGGGTGTCTATGTTTAAGTCTTTAGCTCTTGCTTACAA 135

Qy   104 TGCACTTTTCCGAGGGCTTGACATCAGGGTCTATGTTTAAGTCTTTAGCTCTTGCTTACAA 163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104 TGCACTTTTCCGAGGGCTTGACATCAGGGTCTATGTTTAAGTCTTTAGCTCTTGCTTACAA 163

Qy   136 AGACCACGGCAAATTCCTTCCTCTGAAAGCCCTCCGCAGCCCCCACAGCGCGCCCTCGCGCAGGGCCCCAG 195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   136 AGACCACGGCAAATTCCTTCCTCTGAAAGCCCTCCGCAGCCCCCACAGCGCGCCCTCGCGCAGGGCCCCAG 195

Qy   164 AGACCACGGCAAATTCCTTCCTCTGAAAGCCCTCCGCAGCCCCCACAGCGCGCCCTCGCGCAGGGCCCCAG 223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 AGACCACGGCAAATTCCTTCCTCTGAAAGCCCTCCGCAGCCCCCACAGCGCGCCCTCGCGCAGGGCCCCAG 223

Qy   196 CCTGGCCGCCTACTGCCCAGCAATGCCCTCCAGCCGCGCCCCGGGGGACCACCAGCAGCTCCTCT 255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   196 CCTGGCCGCCTACTGCCCAGCAATGCCCTCCAGCCGCGCCCCGGGGGACCACCAGCAGCTCCTCT 255

Qy   224 CCTGCCGCCTACTGCCCAGCAATGCCCTCCAGCCGCGCCCCGGGGGACCACCAGCAGCTCCTCT 283
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224 CCTGCCGCCTACTGCCCAGCAATGCCCTCCAGCCGCGCCCCGGGGGACCACCAGCAGCTCCTCT 283

Qy   256 CTGGAGCGGGAGGACGGATGATCGAAAGGAAGGAAGGAGAGGAGCAGGAGAACCCGTGGCAAGGAA 315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284 CTGGAGCGGGAGGACGGATGATCGAAAGGAAGGAAGGAGAGGAGCAGGAGAACCCGTGGCAAGGAA 343

Qy   316 GAGCGCCAGGAGGCCCAGCGCCCTGCCGGGAACGCAAG 375
         |||||||||||||||||||||||||||||||||||||||
Db   344 GAGCGCCAGGAGGCCCACGCGCCCTGGCCGGGAACGCAAG 383

Qy   376 CACCCACCGGTGGAAAGCAGTGACACCCCCAAGGACCCAGCAGTCTCAG 435
         |||||||||||||||||||||||||||||||||||||||||||||
Db   404 CACCCACCGGTGGAAAGCAGTGACACCCCCAAGGACCCAGCAGTCTCAG 463

Qy   436 CCCATGGCCCAGGACTCTCTGGCCCCTTCCAGATCTGCTACCCAATGGAGGAGGACTTTGGAGAAGCGG 495
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   464 CCCATGGCCCAGGACTCTCTGGCCCCTTCAGATCTGCTACCCAATGGAGGACTTTGGAGAAGCGG 523

Qy   496 AGTGAACCCCAACCTGAGGAGGGGGAGCCCAGCTGCAGGGCCCCCAGCT 555
         |||||||||||||||||||||||||||||||||||||||
Db   524 AGTGAACCCCAACCTGAGGAGGGGAGCCCAGCTGCAGGGCCCCCAGCT 583

Qy   556 GAAGGAGGAGGAACTGACAGCCCCACCCCACCCCAGCAGCCTGGCTGTGT 615
         |||||||||||||||||||||||||||||||||||||||||||
Db   584 GAAGGAGGAGGAACTGACAGCCCCACCCCAGCAGCCTGGCTGTGT 643

Qy   616 GTGACCAAGGAGGGAACTGACACCCCACCAGAAACCTGGAGGGGAGCCCAGCT 675
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   644 GTGACCAAGGAGGGAACTGACACCCCACCAGAAACCTGGAGGGGAGCCCAGCT 703

Qy   676 ATCGAATCCATGAAAATGGAGGGCTCCCGGGCCCGCACTGCGAGGTGGCTTGGGCTGTGCTGT 735
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   704 ATCGAATCCATGAAAATGGAGGGCTCCCGGGCCCGCACTGCGAGGTGGCTTGGGCTGTGCTGT 763

Qy   736 TCCAGCCTCCGTCCAGCCACCATGCCCAAGACTCACCTTCCAAGGACTCACCCTACTAC 795
         ||||||||||||||||||||||||||||||||||||||||
Db   764 TCCAGCCTCCGTCAGCCACCATGCCCAAGACTCACCTTCCAAGGACTCACCCTACTAC 823

Qy   796 ATCAGCAAAACGGAAAACGGGGGATGAGTGGGCTGGCAGCAGCCTGACAAGAAA 855
         ||||||||||||||||||||||||||||||||||||||||||||
Db   824 ATCAGCAAAACGGAAAACGGGGGATGAGTGGGCTGGCAGCAGCCTGACAAGAAA 883

Qy   856 GCCAAGGTAATTCCAGTAATGAATGCTGTGGAAGAGACAAACCAGGCCTCGGGAGAGTCTCAG 915
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   884 GCCAAGGTAATTGCAGTAATGAAATGCTGTGGAAGAGACAACCAGCCTTGGGAAGAGTCTCAG 943

Qy   916 AAGGTGGAGGAGAAGGCCAGCGCCCCTCCTGGTGTGCCAGGCCACCCGGACCCTGCTTCTCCGACT 975
         ||||||||||||||||||||||||||||||||||||||||||
Db   944 AAGGTGGAGGAGAAGGCCAGCGCCCTCCTCCGCAGCAGCCCTGCTTCTCCGACT 1003

Qy   976 GTGGGCACCACCCCCTGAGCCACCAGTAGGGAGGGGGGATGCTTGGGGGGGACAAGAATGCTACCAAAGCA 1035
         ||||||||||||||||||||||||||||||||||||||
Db 1004 GTGGGCCACCACCCCCTGAGCCACCAGTAGGGAGGGGATGCTGGGGGACAAGAATGCTACCAAAGCA 1063

Qy 1036 GCCGACGATGAGCCTGAGTATGAGGATGGGCCGGGGGGCTTTGGCATTGGAGAGCTGGTGTGTGG 1095
         |||||||||||||||||||||||||||||||||||||||||||||||
Db 1064 GCCGACGATGAGCCTGAGTATGAGGATGGGCCTGGGGGGCTTTGGCATTGGAGAGCTGGTGTGTGG 1123

Qy 1096 GGGAAAACTTCGGGGCTTCTTCCCTGGTGGGCCAGGCCGGAATTGTTGGTGGATCACAGGC 1155
         |||||||||||||||||||||||||||||||||||||||||||
Db 1124 GGGAAAACTTCGGGGCTTCTTCCCTGGTGGGCCAGAATTGTCTCTTCTTGGTGGATCACAGGC 1183

Qy 1156 CGGAGCCGAGCAGCTGAAGGCACTCGCGTGGGGTCATGTGGTTCGGAGATGGCAAGTTCTCA 1215
         |||||||||||||||||||||||||||||||||||||||
Db 1184 CGGAGCCGAGCAGCTGAAGGCACTCGCTGGGTTCGGAGATGGCAAGTTCTCA 1243

Qy 1216 GTGGTGTTGTGTTGGAGAAGCTCATCCGGCTCGAGCTCCTCTTCTCGCAGTGCATTCCACCAGGCC 1275
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 1244 GTGGTGTGTGTTGGAGAAGCTCATGCCGCTCATCGCTTCTCGCAGTGCATTCCACCAGGCC 1303

Qy 1276 ACCTACAACAAGCAGCACCCCATGTACCGCAAAGCCAATCTACGAGTCCTCCAGGTGGCCAGC 1335
         |||||||||||||||||||||||||||||||||||||||||
Db 1304 ACCTACAACAAGCAGCACCCCATGTACCGCAAAGCCAATCTACGAGTCCTCCAGGTGGCCAGC 1363

Qy 1336 AGCCGTCCGGGAGAGCTGTTTTCCAGCTTGCCATGACAGTGATGAAAAGTGACAGTGCCAAG 1395
         ||||||||||||||||||||||||||||||||||||||||
Db 1364 AGCCGTCCGGGAGAGCTGTTTCCAGCTTGCCATGACAGTGATGAAAAGTGACAGTGCCAAG 1423

Qy 1396 GCTGTGGGAAGTGCAGAACAAGCAGAGATGAATTGAAATGGGCCCTTGGGCTTCCAGCCCCTCG 1455
         |||||||||||||||||||||||||||||||||||||||
Db 1424 GCTGTGGGAAGTGCAGAACAAGCAGAGATGAATTGGGCCCTTGGGCTGGCTTTCCAGCCCCTCG 1483

Qy 1456 GGTCCTTAAGGGCCTTGGAGCCGCACCAGAGAAGAGAGAAGAGAAGAGAAATCCTTACACACC 1515
         |||||||||||||||||||||||||||||||||||||||||||
Db 1484 GGTCCTTAAGGGCCTTGGAGCCACCAGAGAAGAGAAGAGAAATCCTTTACACAGAAGTTTTACACC 1543

Qy 1516 GACATGTGGGTGGGAGCCCTGAAAGCACAGCTGCTTACGCCCCCACCCAGCCAAGAAAACCC 1575
         ||||||||||||||||||||||||||||||||||||||||||||
Db 1544 GACATGTGGGTGGGAGCCTGAAAGCACAGCTGCTTACGCCCCCACCCAGCCAAGAAAACCC 1603

Qy 1576 AGAAAGAGCACACAACAGAAGAAAACCTAAGGTCAAGGAGGACATTGATGAGCGCCACAAAGGGAG 1635
         |||||||||||||||||||||||||||||||||||||||
Db 1604 AGAAAGAGCACACAACAGAAGAAAACCTAAGGTCAAGGAGGACATTGATGACCGCCACAAAGGGAG 1663

Qy 1636 CGGCTGGTGTATGAGGTGGCGCCAGAAGTGCAGAAGAAGAAGAAGAAGAAGAGAAGAGAAGACCTCCGGGTGGC 1695
         |||||||||||||||||||||||||||||||||||||||||
Db 1664 CGGCTGGGTGTATGAGGTGGCGCCAGAAGTGCAGAAAACATCGAGAAGTTCCTCGGGTGGC 1723

Qy 1696 GGGAGCCTCAAATGTCACCCTGGGACACCCCACTCTTCATTGGAGGACCAGAACTGT 1755
         |||||||||||||||||||||||||||||||||||||||
Db 1724 GGGAGCCTCAAATGTCACCCTGGGACACCCCACTCTTCATTGGCCAGAACTGT 1783

Qy 1756 AAGAAACTGCTTCTTGGAGTGGTGCTTACCAGTATGACAGATCCTATTGC 1815
         |||||||||||||||||||||||||||||||||||||||||
Db 1784 AAGAAACTGCTTCTTGGAGTGTGCTTACCAGTATGACCAGTCCTATTGC 1843

Qy 1816 ACCATCTGCTCGTGGGGTGGGGGGGTGAAGTGCTCATTGTGTGGGACAACAACTGCTCGAGGTGC 1875
         |||||||||||||||||||||||||||||||||||||||||
Db 1844 ACCATCTGCTCGTGGGGTGGGGGGTGAAGTGCTCATGTGTGGGACAACAACTGCTCGAGGTGC 1903

Qy 1876 TTTTGTGTGTCGGAGTGTGTGGGGATCTCTTGGTGGGGGCCAGGAGCTCAGGCAGCCATTAAG 1935
         |||||||||||||||||||||||||||||||||||||||||
Db 1904 TTTTGTGTCGGAGTGTGTGGGGATCTCTTGGTGGGGCCAGGAGCTGCTCAGGCAGCCATTAAG 1963

Qy 1936 GAAGACCCCTGGAACTGCTTACATGTGCGGGCATAAGGGCAGCACCTATGGCGTCGCGAAGA 1995
         |||||||||||||||||||||||||||||||||||||||||
Db 1964 GAAGACCCCTGGAACTGCTTACATGTGCGGGCATAAGGGCCAGCACCTATGGCGTCGGAAGA 2023

Qy 1996 CGGGAAGACTGGGCCTTCGACTCCAGATGTTCTTTGCCAATAACCATGACCAGGAAATTT 2055
         ||||||||||||||||||||||||||||||||||||||||
Db 2024 CGGGAAGACTGGGCCTTCGACTCCAGATGTTCTTTGCCAATAACCATGACCAGGAAATTT 2083

Qy 2056 GACCCCCCAAAGGTTTACCCCACCTGTGCCACTGAGAGAGAGAGAGAGCCCATCCGGGTGCTG 2115
         ||||||||||||||||||||||||||||||||||||||||
Db 2084 GACCCCCCAAAGGTTTACCCCACCTGTGCCACTGAGAGAGAGAAGCCCATCCGGGTGCTG 2143

Qy 2116 TCTCTCTTGATGGGATTGCTACAGGGGCTCCTGGTGGGCTGAAGGAGACCTGCAAGTG 2175
         |||||||||||||||||||||||||||||||||||||||||
Db 2144 TCTCTCTTTGATGGGATTGCTACAGGGGCTCCTGGTGGGCATCCAAGTG 2203

Qy 2176 GACCGGCTACATTGCCTCCCTCCCAGGTGTGTGAGACTCCATCACACGGTGGGGCATGGTCCGGCAC 2235
         ||||||||||||||||||||||||||||||||||||||||
Db 2204 GACCGGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGGTGGGGCATGGTCCGGCAC 2263

Qy 2236 CAGGGGAAAGAATCAATGTACGTCGGGGGACGTCCGCACGGTCACACAGAGGCCATATCCAGGAG 2295
         ||||||||||||||||||||||||||||||||||||||||
Db 2264 CAGGGGAAAGAATCAATGTACGTCGGGGACGTCCGCAGGTCACACAGAGGCCATATCCAGGAG 2323

Qy 2296 TGGGGCCCATTCACACCTGCTGATTGGATGGCAGTCCCTGCCAAATGACCCTTCCATTGTCAAC 2355
         |||||||||||||||||||||||||||||||||||||||
Db 2324 TGGGGCCCCATTCACACCTGTGTGATTGGATGGCAGTCCCTGCGAATGACCCTTCCATTGTCAAC 2383

Qy 2356 CCTGGCCCGCCAAGGGACTTTATGAGGGTACTGGCGCCGGCCTGGGTGTGTGAGGATTCTACCGCCTC 2415
         ||||||||||||||||||||||||||||||||||||||||
Db 2384 CCTGGCCCGCCAAGGGACTTTATGAGGGTACTGGCGCCCTCTGTTCTCTACCGGCCC 2443

Qy 2416 CTGCATGATGCGCGGGGCCCCAAAGAGAGGAGAGATCATCGCGCCCCTTCTTCGCTCTTTGGAGAAT 2475
         |||||||||||||||||||||||||||||||||||||||||
Db 2444 CTGCATGATGCGGCGGGCCCAAGGAGAGGAGAGATCGGCGCCCCTTCTTCTCTTTGAGAAT 2503

Qy 2476 GTGGTGGGCCATGGGCGTTAGTGACAAGAGGGGACACATCTCGCGCGATTCTTGAGTCTAACCCC 2535
         ||||||||||||||||||||||||||||||||||||||||
Db 2476 GTGGTGGGCCATGGGCGTTAGTGACAAGAGGGGACACATCTCGCGCGATTTCTTGAGTCTAACCCC 2535

```
Db  2504 GTGGTGGCCCATGGGCGGCGTTAGTGACGAAGAGGGACATCTCGGCGATTTCTTTTGGGTGGGTAAACCC 2563
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2536 GTGGATGATTGACGCGCCAAAGAAGAGTGTCTCGTGCACACAGGAAACCTTGAGTCTAAACCC         2595

Db  2564 GTGATGATTGACGCGCCAAAGAAGAGTGTCTCCTGCACACAGGGCCCGTTACTTCTGGGGTAAC        2623
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2596 CTTCCTGGCATGAAACAGGCCCTTTGGCATCCACTGTGAAGATAAGCTGCTGCAAGAG              2655

Db  2624 CTTCCTGGCATGAAACAGGCCCTTTGGCATCCACTGTGAAGATAAGCTGGAGCTGCAAGAG           2683
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2656 TGTCTGGAGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGACCAATTACCACCAGTCA            2715

Db  2684 TGTCTGGAGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGACCAATTACCACCAGGTCA           2743
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2716 AACTCTATAAAGCAGGGCAAAAGACCACGCAGCATTTCCCCGTCTTCATGAACGAAGGAGGAC         2775

Db  2744 AACTCTATAAAGCAGGGCAAAAGACCACGCAGCATTTCCCCGTCTTCATGAACGAAGGAGGAC         2803
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2776 ATCCTGTGGTGGTGCACTGAAAATGGAAAAGGGGTGTTTGGGTGTTTGGGCTTCCCGTC            2835

Db  2804 ATCCTGTGGTGGTGCACTGAAAATGGAAAAGGGGTGTTTGGGTGTTTGGGCTTCCCGTC            2863
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2836 TCCAACATGAGCGCCGCTTGGGCGAGGACAGAGACTGCTCTGGGCGCCGATCGTGGGAGG           2895

Db  2864 TCCAACATGAGCGCCGCTTGGGCGAGGACAGAGACTGCTCTGGGCGCCGATCGTGGGAGG           2923
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2896 ATCCGCCACCTCTTCGCGCCTGAAAGGAAAATATTTTGGCTTGTGTGTGAAGGACATGGGGGG        2955

Db  2924 ATCCGCCACCTCTTCGCGCTCCGGCTGAAAGGAATATATTTTGGCTTGTGTGTGAAGGACATGGGGG     2983
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2956 CAAACTGAAGCTAGTGATGATAAAAAAAGTTAAACAAAACAAAAAACAAAAAAACAAA             3015

Db  2984 CAAACTGAAGTAGTGATGATAAAAAAAGTTAAACAAAACAAAAAACAAAAAAACAAA             3043
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3016 ACAATAAAAACACCAAGAACGAGAGGACGGAGAAAAAGTTCAGCACCGAGAGAGAAGAAAAGG        3075

Db  3044 ACAATAAAAACACCAAGAACGAGAGGACGGAGAAAAGTTCAGCACCCAGAGAGAAAAAAGG          3103
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3076 AATTTAAAGCAAAACCACAGAGAGGAGGAAAAACGCCGGAGAGGGGCTTGGCCTTGCAAAAAGGGGTTG  3135

Db  3104 AATTTAAAGCAAACCACAGAGAGGAGGAAAAACGCCGGAGGGGCTTGGCCTTGCAAAAGGGGTTG      3163
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3136 GACATCATCTCCTGGAGTTTTCAATGTTAAACCTTCAGTCCTATCTAAAAAAGCAAAAATAGGC       3195

Db  3164 GACATCATCTCCTGAGTTTTCAATGTTAAACCTTCAGTCCTATCTAAAAAAGCAAAAATAGGC        3223
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3196 CCCTCCCCCTTCTTCCCCTCCGGTCCTAGGAGGCGGAACTTTTGGTTTTTTGTTTTTTCTACTCTTTTTCA 3255

Db  3224 CCCTCCCCCTTCTTCCCCTCCGGTCCTAGGAGGCGGAACTTTTTTGTTTTTCTACTCTTTTTCA       3283
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3256 GAGGGGGTTTTCTCGTTTGTTGTTTTGTTTGGGTTTGGGTTTTTTGTTTGTTGTTTGTTGTTAT       3315

Db  3284 GAGGGGTTTTCTCGTTTGTTGTTTGGGTTTTTTTGTTTGTTGTTGACTGGAAAACAAGAGAGTTAT     3343
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3316 TGCAGCAAAATCAGTAACAAACAAAAAAGTAGAAATGCCTTGGAGAGGAGAAAAAGGGAGAGGAGG     3375

Db  3344 TGCAGCAAAATCAGTAACAAACAAAAAAGTAGTA--GCCTTGGAGAGGAGAAAAAGGGGAGAGGAGG    3401
         |||||||||||||||||||||||||||||||||      ||||||||||||||||||||||
Qy  3376 AAAATTCTATAAAAAACTTAAAATATTGGTTTTTTTTTTTTTTTTCCTTTTCTATATATATCTCT     3435

Db  3402 AAAATTCTATAAAAAACTTAAAAATATTGG-TTTTTTTTTTTTTTTTCCTTTTCTATATATCTCT     3460
         |||||||||||||||||||||||||||||       ||||||||||||||||||||
Qy  3436 TTGGTTGTCTCTAGCCTCATCAGATAGGAGCACACAAACAGGGAGGGAGAATAATAGAGACCCCTCGG  3495

Db  3461 TTGGTTGTCTCTAGCCTCGATCAGATAGGAGCACAAACAGGGAGGAATAGAGACCCTCGG          3520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3496 AGGCAGAGTCTCTCCTCCTCCCA-CCCCCGAGCAGCGAGTCTCAACAGCACCACCATTCCTGTCGTCATGCA 3555

Db  3521 AGGCAGAGTCTCTCCTCCTCCCCA-CCCCCGAGCAGTCTCAACAGCACCACCATTCCTGTCATGCA    3579
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3556 AAACAGAACCCAACTAGCAGCAGGGCGGCTGAGCAGAAGACACCACACCAGAGACAC-TTCTACA     3614


Db  3580 AAACAGAACCCAACTAGCACCAGGGCGCCGGCTGAGAGGAGAGAACACCACCAGAGACACTTTTCTACA 3639
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3615 GTATTTCAGGTGCCTACCACACAGGAAAACCTTGAAGAAAAAACCAGTTTCTAGAAGCCGCTG        3674

Db  3640 GTATTTCAGGTGCCTACCACACAGGAAAAACCTTGAAGAAAAAACCAGTTTCTAGAAGCCGCTG       3699
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3675 TTACCTCTTGTTTACAGTTTATATATATATATGATGATGATGATGATGATATATATATATAAAAGG    3734

Db  3700 TTACCTCTTGTTTACAGTTTATATATATATGATGATGATATATATATATATATATATAAAAGG        3759
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3735 TACTGTTAACTTACTGTTACATCCCCGACTTCATCATGTGTGCTTTCAAAAACAGCGAGGAGATGAGC  3794

Db  3760 TACTGTTAACTTACTGTTACATCCCCGACTTCATCATGTGTGCTTTCAAAAACAGCGAGGAGATGAGC  3819
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3795 AAAGACATCAGCTTCCCGCCCTGGCCCCTCTGTGCCAAAGGGTTGCACCCAGGATGGGGACG        3854

Db  3820 AAAGACATCAGCTTCCCGCCCTGGCCCCTCTGTGCCAAAGGGTTGCACCCAGGATGGGGACG         3879
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3855 GGGAGCAGCTGGAGGGGGGTTTAACAAACTGAAGGATGACCCATATATCACCCCCCACCCCT        3914

Db  3880 GGGAGCAGCTGGAGGGGGGTTTAACAAACTGAAGGATGACCCATATATCACCCCCCACCCCT        3939
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3915 GCCCCATGCCTAGCTTCACCTGCCAAAAAAGGGGCTCAGCTGAGGTGGTCGGGACCCTGGGG       3974

Db  3940 GCCCCATGCCTAGCTTCACCTGCCAAAAAAGGGGCTCAGCTGAGGTGGTCGGGACCCTGGGG       3999
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3975 AAGCTGAGTGTGGAATTTATCCAGACTCGCGTGCAAATAAACCTTAGAATATGAAATCTAAAA      4034

Db  4000 AAGCTGAGTGTGGAATTTATCCAGACTCGCGTGCAAATAAACCTTAGAATATGAAATCTAAAA      4059
         |||||||||||||||||||||||||||||||
Qy  4035 TGACTGCCTCAGAAAAATGGCTTGAGAAAACA                                    4066

Db  4060 TGACTGCCTCAGAAAAATGGCTTGAGAAAACA.AAAAAAAAAAAAAAAAA.           4091
```

RESULT 3

AF331856

```
LOCUS       AF331856                4258 bp    mRNA    linear   PRI 02-JAN-2002
DEFINITION  Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,
            complete cds.
ACCESSION   AF331856
VERSION     AF331856.1  GI:18033252
KEYWORDS    .
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4258)
  AUTHORS   Ni,J., Pradhan,S. and Roberts,R.J.
  TITLE     Cloning, expression and characterization of human DNMT3 genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4258)
  AUTHORS   Ni,J., Pradhan,S. and Roberts,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
            Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1..4258
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     gene            1..4258
                     /gene="DNMT3A"
     CDS             227..2956
                     /gene="DNMT3A"
                     /codon_start=1
                     /product="DNA cytosine methyltransferase 3 alpha"
                     /protein_id="AAL57039.1"
                     /db_xref="GI:18033253"
                     /translation="MPSSGPGDTSSAAEREEDRKDGEEQEEPRGKEEROEPSTTARK
                     VGRPGRKRKHPVESGDTPKDPAVISKSPSMAQDSGASELLPNDDLEKRSEPQPEEGS
                     PAGGQKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIESMKM
                     EGSRGRLRGGLGWESSLRQRPMRLTFQAGDPYYISKRKRDEWLARWKREAEKKAKVI
```

```
AGMNAVEENQGPGESQKVEEASPPAVQQPTDPASPTVATTPEPVGSDAGDKNATKAGD
DEPEYEDGRGFGIGELVWGKLRGFSWWMPGRIVSWWMTGRSRAAETRWVHMPGDGKFS
VVCVEKLMPLSSFCSARFHQATNRQPMYRKAIYEVLQVASSRAGKLFPVCHDSEDSDT
AKAVEVQNKPMIEWALGGFQPSPSGPKGLEPEEEKNPYKEYTDMWVEPEAAAYAPPPP
AKRPKRSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCSSLNVTLEHPLFVG
GMCQNCKNCFLECAYQYDDGYQSYCTICCGGREVLMCGNNNCCRCFCVECVDLLVGP
GAAQNAIKEDFWNCTWCGHKGTYGLLRREDWPSRLQMFFANNHDQEFDPPRKVIPPVP
AEKKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVG
DVRSVTQKHIQEWGPPDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLHDARP
KEGDDRPFFWLFENNVAMGVSDKRDISREFLESNPVMIDAKEVSAARHARYFWGNLPGM
NRPLASTVNDKLELQCLEHGRIAFKSKVRTITTRSNSIKQGKDQHFPVFMNEKEDIL
WCTEMERVFGFPVHYIDVSNMSSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV"
```

Query Match   68.6%;   Score 2877.4;   DB 9;   Length 4258;

Best Local Similarity   84.3%;   Pred. No. 0;

Matches 3543;   Conservative   0;   Mismatches   562;   Indels   99;   Gaps   23;

BASE COUNT   1062 a   1089 c   1257 g   850 t

ORIGIN

```
Qy   1979 ATGGGCTGCCTGCGGAAGACGCGGAAGACACTGGGCCTTCTCCGACTCCAGATGTTCTTTGGCCAATA 2038
Db   1992 ACGGGCTGCCTGGGCGCGGAGAGGACTGGGCCCTCCCCAGATGTTCTTCGCTAATA 2051

Qy   2039 ACCATGACCAGGAAATTTGACCCCCCGCCAAAGGTTTACCCGCTGTGCCAGCTGCTGAGGAAGAAGGA 2098
Db   2052 ACCACGACCAGGAATTTGACCCTCCAAAGGTTTACCCGACCGTGCCCAGCTGGAGAGAGGGGA 2111

Qy   2099 AGCCCATCCGCGTGCGTGCTGTCTCTCTTTCGATGGGATTGGCTACACAGGGCTCCTGGTTGCTGAAGG 2158
Db   2112 AGCCCATCCGGGGTTTGGCGTGTCTCTCTTTGAATGGAAATCGGCTCCCTACACAGGGCTTCTGAAGG 2171

Qy   2159 ACCTGGGCCATCCAAAGTGGACCGGCCATCCAGGTTGGGACACGCTACACATTTGGCCTCGGAGGGCA 2218
Db   2172 ACTTGGGCCATTCAGGTTGGCACCGCTACATTTGGCCTCGGAGGGGACTCACACACGG 2231

Qy   2219 TGGGCATGGTGCGGGCACCCAGGGAAAAAGATCATGTACGTGCCTCCCGCAGGGCTCACAC 2278
Db   2232 TGGGCATGGTGCGGGCACCCAGGGAAAGATCCATGTACGTGCGGGGACGCGTCCCGCAGGCGTCACAC 2291

Qy   2279 AGAAGCATATCCAGGAGTGGGGGCCCATTCGACCTGGTGATTGGAGGCCAGTCCCTGCCAAATG 2338
Db   2292 AGAAAGCATATCCAGGAGTGGGGGCCCATTCGATCTGGTTAGTGCACACAGGGACAATCTCGGCGAT 2351

Qy   2339 ACCTCTCCCATTGTCAACCCCGTGACGGCCAAAGGAGACTTTATGAGGGTACTGGCGCGCCCCTCTTCT 2398
Db   2352 ACCTCTCCATGTGTTCAACCCTGCTCGCAAAGGGCCCTCTACGAGGGGACTCACACAGGG 2411

Qy   2399 TTGAGTTCTACCGGCCTCCTGCATGATGCCGCGCCCCAAGGAGAGGGAGGAGATGAATCGCGCCCCTTCT 2458
Db   2412 TTGAGTTCTACCGCCTCCTGCATGATGCCGCGCCCCAAGGAGAGGGAGGAGATGATCGGCCCCTTCT 2471

Qy   2459 TCTGGGCTCTTTGAGAATGTGTGGGGCATGGGGCGTTAGTGCAAGAGGGACATCTCGCGAT 2518
Db   2472 TCTGGGCTCTTTGAGAATGTGGGTGGGCCATGGGCGTTAGTGCACAGAGGGACATCTCGGCGAT 2531

Qy   2519 TTCTTGAGTCTAAACCCCGTGATGAATGTGCACGGCCAAAGAGAAGTGTCTCTGCACACAGGGCCC 2578
Db   2532 TTCTCGAGTCCAAACCCTGTGATGAATGTGCACGGCCAAAGAGTTCACGTGCACACAGGGCCC 2591

Qy   2579 GTTACTTCTGGGGTAACCTTTCCTGGGTTTGGCATCCACCCACGTGTGAATGATA 2638
Db   2592 GCTACTTCTGGGGTAACCTTCCCGGTTATGAACAGGCCGTTGGCATCCACCACTGTGAATGATA 2651

Qy   2639 AGTTGGAGCTGCAAGAGTGTCTGGGAGGCACGGCGAGAATAGCCCAAGTTCAGCAAGTGAGGA 2698
Db   2652 AGCTGGAGCTGCGAGGAGTGCTGGGACCATGGCCAGGATAGCCAAGTTCAGCAAGTGAGGA 2711

Qy   2699 CCAATTACCACCAGGTCAAAACTCTATAAAGCAGGGGCAAAAGACCACAGCATTTCCCCGTCTTCA 2758
Db   2712 CCATTACTACCAGGTCAAAACTCCATAAAAGCAGGGGCAAAAGACCACAGCATTTTCCCCGTCTTTCA 2771

Qy   2759 TGAACGAGAGGAGGAGGACATCCTGTGGTGCACTGAAAATGGAAAGGGGTGTTGGCTTCCCCG 2818
Db   2772 TGAATGAGAAAAGAGGACATCTTATGGTGCACTGAAATGGAAAGGGGTATTTGGTTTTCCCAG 2831

Qy   2819 TCCACTCACACAGACAGTCTCCAACATGAGCCGCCTTGGCGAGCCGAGGACTCGGGCCGGAT 2878
Db   2832 TCCACTATACTCACGTCTCCAACATGAGCCGCCTTGGCAGGCAGAGACTCGCTGGGCGGGT 2891

Qy   2879 CGTGGAGCGTGCCGTGGTCATCCGGCCACCTCTTCGCTCCGCGTTGAAAGGAAATATTTTGCTTGTG 2938
Db   2892 CATGGAGCGTGCCAGTCATCCGCACTCTTTCGGCTCCGCTCAAGGAGTATTTTTGGGTGTTG 2951

Qy   2939 TGTAAGGGACATGGGGCCAAAACTGAAGTAGTGATGATAAAAAGTTAAACAAAACAAAACAA 2998
Db   2952 TGTPAAGGGACATGGGGGCCAAAACTGAGGTAGGGA----CACAAAGTTAAACAAAACAAAC-- 3005

Qy   2999 ACAAAAAAACAAAAACAATAAAAACACACAAAGAACACGAAGGAGGACGGGAGAAAAGT-TCAGC 3057
Db   3006 ---AAAAAACACAAAACATAATAAAAACACCAAGAACATGAGACGGATGGAGAGAAGTATCAGC 3062
```

```
Qy   3058 ACCCAGAAGAAGAAGAAGAAGGAAATTTAAAAGC--AAACCACAGAGAGGAGGAGAAAAAGCCCGGAGGGC 3115
Db   3063 ACCCAGAAGAGAAGAAAAAAGGAAATTTAAAACACAGAGAGGAGGAGAAATACCGGAGGGGCT 3122

Qy   3116 TTGGCCTTGCCAAAAGGGGTTGGACATCATCTCCTGAGTTTTCAATGTTAACCTTCAGTCCT 3175
Db   3123 TTGGCCTTGGGAAAAGGGTTGGACATCANTCTCCTGAATTTTTCAAATGTTATTCTTCAGTCCT 3182

Qy   3176 ATCTAAAAAGCAAAATAGGC-CCCTCCCCTTCTTCCCC---TCCGGTCCTAGGAGCGCGAA 3231
Db   3183 ATTTAAAAAACAAAACCAAGCTCCCCTTCCCCTTCTCCCCTTCAAATGTTCTTTCAGTCCT 3242

Qy   3232 CTTTTTATTTTTCTACTCTCTTTTTTCAGAGGGGGGTTTTCTCGTTTGGGTTTTTTTTGTTTGT 3291
Db   3243 CCTTTTTATTTTTCTACTCTT-TTTCAGAGAGGGTTTTCTGTTTGTTTGTTGGTTTTTTTTTTCGTTG 3300

Qy   3292 CTGTGACTGAAACAAAGAGAGTTATTGCAGC-AAAATCAGTAACAACAAAAAGTAGAAATG 3350
Db   3301 CTGTGACTGAAACAAGAAAGGTTATTGCAGCAAAAATCAGTAACAAAAATAGTAACAATA 3360

Qy   3351 CCTTGGAGAGGAAAGG------GAGAGAGGAGAAATTCTATAAAAAACTTAAAATATT 3401
Db   3361 CCTTGCAGAGGAAAGGTGGGAGAGAGGAGAAAAAGAAATTCTATAGAAATCTATATATTTG 3420

Qy   3402 GGTTTTTTTTTTTTTTTTTTTCCTTTT--------TCTATATA 3430
Db   3421 GGTTGTTTTTTTTTTTTTGTTTTTTTGTTTTGGGTTTTTTTTTACTATATATC 3480

Qy   3431 TCTCTTTTGGTTGTTGTCTCTAGCCCTGTAGTCAGATAGGACGGACACAACA--GGAAAGAGAATAGAGA 3488
Db   3481 TTTTTTTTGCTTTGTCTCTAGCCCTGATCAGATAGGACGGACAACCTTGAAGAAAATCAGTTTCTAGAA 3540

Qy   3489 CCCTCGGAGGCAG--AGTCTCCTCTCCCACCCCCGAGCAGTCTCAACAGCACCATTCCTG 3547
Db   3541 GACACTCAGGCGGCCAGCATTCCCTCCCCAGCCACTGAGCTGTCGTGCCAGCACCATTCCTG 3600

Qy   3548 GTCATGCAAAAACAGAAACCCAACTAGCAGCAGCGGCCGGTGAGAAGAAACACCACCAGACACT 3607
Db   3601 GTCACGCAAAACAGAAACCCAGTTAGCAGCAGGGAGAGGAGAACACCACCAAGACA-TTT 3659

Qy   3608 TTCTACAGTATTTCAGGGTCCCTACCACACACAGGAAGAAAACCCAAGTTTGTAGAA 3667
Db   3660 TTCTCACAGTATTTCAGGTGCCTACCACCACACAGGAAAACCTTGAAGAAAATCAGTTTCTAGAA 3719

Qy   3668 GCCCCTGTTACCCTCTTGTTTACAGTTTATATATATATGATAGAATATGAGATAATATATA 3727
Db   3720 GCCGGCTGTTACCCTCTTGTTTACAGTTTATATATATGATAGAATATGAGAT--ATATATATA 3777

Qy   3728 TAAAAGGTACTGTTTAACTACTGTACATCCCGAACTTCAATAATGGTGGTCTTTCAAAAACAGCGGA 3787
Db   3778 TAAAAGGTACTGTTTAACTACTGTACAACCCGACTTCAATAATGGTGGTCTTTC-AAACAGCGGA 3836

Qy   3788 GATGAGCAAAGACACATCAGCTTTCCGCCCTGGCCCTTCTGTTGCAAAGGGGTTTCAGCCCCAGGATG 3847
Db   3837 GATGAGTAAAAACATCAGCTTTCCACGTTGCCCTTTCTGCGGCAAAAGGGTTTCACCCAAGGATGG 3896

Qy   3848 GGGAGAGGGGGACCAGCCTGGAGGGGGGGTTTTAACAAAAACTGAAGGATGACCCATAATCACCCCC 3907
Db   3897 AGAAAGGGAGACAGCTTCCAGATGCGGGGTTCTACCGGTGGGCTCTTCACGGTGGTGTT 3956

Qy   3908 CAC-------CCCTGCCCATGCCTAGCTTCACCTCCCAAAAAAGGGGTCAGCT 3954
Db   3957 AACGAAAGTCAAGGAGGAGAACTGGGAGCCAGGTTCTCCCTGCCCAAAAAGGGGGGCTAGAT 4016

Qy   3955 GAGGTGGGTCGGGACCCTGGGGAAAGCCTGAGGTGTGGAAATTTATCCCGGACTCGCCTGCAATAAC 4014
Db   4017 GAGGTGGGTCGGGGCCCGTGGACAACGTGAGAGTGGGACCAGTTTATCCCGGCTCGCCTGCAATAAC 4076

Qy   4015 CTTAGAAATATGAATCTAAAATGACTGCCTCAGAAAAAAATGGC------TTGAGAAAAACATT 4068
Db   4077 TTTGATTGTTTTTTCTAAAGGGAGGAGACTCCCTCGGCCAAGGGTACGGAGGGTTGAGGGGGTCTTCA 4136

Qy   4069 GTCCCTGATTTTTGTCAGCCACGTTGAAGGCCCCTTGTGTGGGATCAGAAAATATTCC 4128
```

```
Db   4137 GGCCCAGTTTCTCACTTTAGCCAAATTCGAGGGGTCTCCTTGTGGTGGGATCAGAACTAATCC 4196
          | |||   |   |    | ||||   | ||  || ||||      |||||  ||| |  |||
Qy   4129 AGAGTGAGGAAAGTGACCCGCCCATTAACCCNCCTGGAGCAAATAAAAAAACATACAAA 4188

Db   4197 AGAGTGTGGGAAAGTGACAGTCAA--AACCCACCTGGAGCAAATAAAAAAACATACAAA 4254
          |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Qy   4189 AGAGTGTGGGAAAGTGACAGTCAA--AACCCACCTGGAGCAAATAAAAAAACATACAAA 4254

Qy   4189 ATGT 4192
Db   4255 ACGT 4258
```

RESULT 4

```
LOCUS       AF067972                3005 bp    mRNA    linear   PRI 12-FEB-2001
DEFINITION  Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,
            complete cds.
ACCESSION   AF067972
VERSION     AF067972.2  GI:12746531
KEYWORDS    .
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3005)
  AUTHORS   Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.
            and Li,E.
  TITLE     Cloning, expression and chromosome locations of the human DNMT3
            gene family
  JOURNAL   Gene 236 (1), 87-95 (1999)
  MEDLINE   99365304
  PUBMED    10433969
REFERENCE   2  (bases 1 to 3005)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th
            Street, Charlestown, MA 02129, USA
REFERENCE   3  (bases 1 to 3005)
  AUTHORS   Hata,K., Shirohzu,H., Sasaki,H. and En,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th
            Street, Charlestown, MA 02129, USA
  REMARK    Sequence update by submitter
COMMENT     On Feb 12, 2001 this sequence version replaced gi:4927369.
FEATURES             Location/Qualifiers
     source          1..3005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p23; between D2S171 and D2S174; 48-50 cM"
     gene            1..3005
                     /gene="DNMT3A"
     CDS             230..2968
                     /gene="DNMT3A"
                     /codon_start=1
                     /function="de novo DNA cytosine methyltransferase"
                     /product="DNA cytosine methyltransferase 3 alpha"
                     /protein_id="AAD33084.2"
                     /db_xref="GI:12746532"
                     /translation="MPAMPSSGPGDTSSSAAEREDRKDGEEQEEPRGKEREQEPSTT
                     ARKVGRPGRKRKHPPVESGDTKPKDPAVISKSPSMAQDSGASELLPNGDLEKRSEPQPE
                     EGSPAGGQKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIES
                     MKMEGSRGRLRGLGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKA
                     KVIAGMNAVEENQGPEGSQKVEEASPPAVQQPTDPASPTVATTPEPVGSDAGDKNATK
                     AGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDG
                     KFSVVCVEKLMPLSSFSSCSSAFHQATYNKQPMYRAIYEVLCVAVSSRAGKLFPVCHDSDE
                     SDTAXAVEVQNKPMIEWALGGFVQPSGPKGLEPPEEKNPYKEVYTDMWVEPEAAAYAP
                     PPPAAKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVTLEHPL
                     FVGGMCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRFCVCVDLL
                     VGPGAQAAIKEDPWNCYMCGHGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYP
                     PVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM
                     YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLIYEGTGRLFFEYYRLLHD
                     ARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL
```

```
BASE COUNT      674 a    846 c    978 g    507 t
ORIGIN

Qy Query Match 57.1%;  Score 2394.6;  DB 9;  Length 3005;
   Best Local Similarity 89.7%;  Pred. No. 0;
   Matches 2614;  Conservative 0;  Mismatches 287;  Indels 12;  Gaps 4;
```

```
Qy    69 CTGAGGCTGCACTTTTTCCGAGGGGCTTTGACATCAGGGTCTATGTTTAAGTCTTAGTCTTTG 128
         |||||| |   |         |||| || |  |  |  || |||||| |||||||||||||||
Db    86 CAGAGCGCACCGTTCACCAGAGGGCTCAACACCGAGGGATCTATGTTTAAGTTTTAAACTCTCG 145

Qy   129 CTTACAAAGACCACACGGCAAATTCCTTCTTCTGAAGCCCTTCCGAGCCCCACACAGCCCTCGCA 188
         ||  |||||||   || |  ||||||||||||||||||||||||||||||||||||||||||||
Db   146 CCTCCAAAGACCCACGATAATTCCTTCCCCAAAGCCC-AGCAGCCCCCAGC-CCGCCCA 203

Qy   189 GCCCCAGCCCTGCCCGCTTCACTGCCCTACTGCCCA------------GCAATGCCCTCCCACGGC 241
         ||||||||||||| |||  ||||||||||||||||  |||||||||| | |||||||||||||||
Db   204 GCCCCAGCCTGCCTCCCCGGCGCCCCCAGCAGATGCCCGCCCCCAGCCCCGGGGGACA 263

Qy   242 CCAGCAGCTCCTCTTCTGGAGCGGGAGGGAGGGGAGGGATGAAAGGAAGGAGGAGGAACAGGAGGA 301
         |||||||   |||| | |  || |||||||||||||||||||||||||||||||||||||||||||
Db   264 CCAGCAGCTCTGCTGCGGAGCGGGGAGGGAGGGGAGCCCGAAAAGGACGGAGGAGGAGGAGGAGC 323

Qy   302 ACCGTGGCAAGGAAGAGAGAGGGGCCCAGAGTCGGAAGCAGTGAAAAGGAAGGTGGGGAGGGCTG 361
         |||||||||| | | | ||||||||||||| | |||  ||||||||||||||||||||||||||||
Db   324 GCGGTGGCAAGGAGGAGGAGGAGGCCAAGACGAGGAGGAGGAGGAGAGGTGGGCGGCCTG 383

Qy   362 GCCGGAAGCGCAAGCACCCCACCCGTGGAAAGCAGTGACACCCCCAAGGACCCCAGCAGTGA 421
         ||||||||||||||||| || |   |   |||||||||||| |||||||||||||||||||
Db   384 GGAGGAAGCGCAAGCACCCCCCCGTGGAAAGCGGTTGACACGCCCAAAGGACCCTGCGGTGA 443

Qy   422 CCACCAAAGTCTCAGCCCCCATGGCCCCAGAGCCTCTGGCCCCTCAGATTCTCTACCCCAATGGAG 481
         ||||||||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||
Db   444 TCTCCAAGTCCCCCATGGCACCCCCAGGAGCCCTCAGGGCTATTACCCAATGGGG 503

Qy   482 ACTTGGAGAAGCCGGAGTGAACCCCAACCCTGAGGAGAAGCCCCAGCCTGCAGGCAGAGG 541
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   504 ACTTGGAGAAGCGGAGTGAGCCCCAGCCAGCCGGGGAGGCCCCTGGGCGGGGAGG 563

Qy   542 GTGGGGCCCCCAGCTGAAAGGAGAGGG----AACTGAGACCCCACCAGAAGCCCCTCAGAGCTG 598
         ||||||||||||||||||||||||||| || |||||||||||||||||||||||||||||||||
Db   564 GCGGGGCCCCAGCAGAGGGGAGAGGGTGCACTGAGACCTGCGGCCCTGAAGCCTCAAGAGCAG 623

Qy   599 TGGAGAATGGCTCGCTGTGTGACCAAGGAAGGCCCGTTGGAGGGGCCTTCGCAGGAGGGCAAAG 658
         ||   ||   ||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   624 TGGAAAATGGCTGCTGCACCCCAAGGAGGGCCCGACGGAGGCCCTGCACGTTGCACGTGCAAAG 683

Qy   659 AACAGAAAGCAGACCAACATCGAATCCATGAAAATGGAGGGGTC 718
         |||||||||||||||||||||||||||||||||||||||||||
Db   684 AACAGAAGGAGACCAACATCGAATCCATGAAAATGGAGGGGTCG 743

Qy   719 GTGGGCTTGGGCTGGGAGTCCAGCCTCCGTCCGTCGCAGACTTCACCTTCCAGG 778
         |||||| |||||| |||||||||||||||||||||||||||||| |||||||||
Db   744 GTGGCTTGGGCTGGGAGTCCAGCCTCCGTCCGTCGCAGACTTCACCTTCCAGG 803

Qy   779 CAGGGGACACCCCTACTACATCAGCAAACGGAAAGCGGGATGAGTGGCTTGCACCGTTGGAAAA 838
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   804 CGGGGGACACCCTACTACATCAGCAAGCGCAAGCGCGACGAGTGGCTGGCACGGTTGCAAAG 863

Qy   839 GGGAGGGCTGAGAGAAAAGCCAAGGTAATTGCAGTAATGCGTGTGGAGAAGAGAAACCAGG 898
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   864 GGGAGGGCTGAGAAGAAAGCCAAGGTCATTGCGGGAATGAATGCTGTGGAAGAGAAGCAGG 923

Qy   899 CCTCTGAGAGTCTCAGAAGGTGAGAGGGCCGAGCCCCTTGCTGCCACCCCACGG 958
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   924 GGCCCCGGGGAGTCTCAGAAGGTGGAGGAGGCCAGCCCTCCTCCTGCAGCCACCCCACTG 983

Qy   959 ACCCTTGCTTCTCCGGACTGTTGGGCACCACCCCTGAGCCAGTAGGAGGGGAGAGCACTG 1018
         |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   984 ACCCCGCATCCCCCACTGTGGCTACCACGCCTGAGCCCGTTGGGTCCGGATGCTGGGGACA 1043
```

```
Qy        1019 AGAATGCTACCAAAGCAGCCGACGATGAGCCTGAGTATGAGGATGGCCGGGGCTTTGGCA 1078
               ||||||| |||||||||| ||| || ||||| ||||| ||||| ||||||||||||||||
Db        1044 AGAATGCCACCAAAGCAGGCGATGACGAGCCAGAGTACGAGGACGGCCGGGGCTTTGGCA 1103

Qy        1079 TTGGAGAGCTGGTGTGGGGGAAACTTCGGGGCTTCTCCTGGTGGCCAGGCCGAATTGTGT 1138
               |||| |||||||||||||||||||| |||||||||||||||||||||||||| |||||||
Db        1104 TTGGGGAGCTGGTGTGGGGGAAACTGCGGGGCTTCTCCTGGTGGCCAGGCCGCATTGTGT 1163

Qy        1139 CTTGGTGGATGACAGGCCGGAGCCGAGCAGCTGAAGGCACTCGCTGGGTCATGTGGTTCG 1198
               ||||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||
Db        1164 CTTGGTGGATGACGGGCCGGAGCCGAGCAGCTGAAGGCACCCGCTGGGTCATGTGGTTCG 1223

Qy        1199 GAGATGGCAAGTTCTCAGTGGTGTGTGTGGAGAAGCTCATGCCGCTGAGCTCCTTCTGCA 1258
               |||| ||||| ||||||||||||||||| ||||||||| ||||||||||||||| || ||||
Db        1224 GAGACGGCAAATTCTCAGTGGTGTGTGTTGAGAAGCTGATGCCGCTGAGCTCGTTTTGCA 1283

Qy        1259 GTGCATTCCACCAGGCCACCTACAACAAGCAGCCCATGTACCGCAAAGCCATCTACGAAG 1318
               |||| |||||||||||||| ||||||||||||||||||||||||||||||||||||| |
Db        1284 GTGCGTTCCACCAGGCCACGTACAACAAGCAGCCCATGTACCGCAAAGCCATCTACGAGG 1343

Qy        1319 TCCTCCAGGTGGCCAGCAGCCGTGCCGGGAAGCTGTTTCCAGCTTGCCATGACAGTGATG 1378
               |||| |||||||||||||||| || ||||||||||| || |  ||||| ||||| ||||
Db        1344 TCCTGCAGGTGGCCAGCAGCCGCGCGGGGAAGCTGTTCCCGGTGTGCCACGACAGCGATG 1403

Qy        1379 AAAGTGACAGTGGCAAGGCTGTGGAAGTGCAGAACAAGCAGATGATTGAATGGGCCCTCG 1438
               | ||||||| || |||||| ||||| ||||||||||||||  ||||||||||||||||| |
Db        1404 AGAGTGACACTGCCAAGGCCGTGGAGGTGCAGAACAAGCCCATGATTGAATGGGCCCTGG 1463

Qy        1439 GTGGCTTCCAGCCCTCGGGTCCTAAGGGCCTGGAGCCACCAGAAGAAGAGAAGAATCCTT 1498
               | ||||||||||||| || || |||||||||||||||||||||||||||||||||||| |
Db        1464 GGGGCTTCCAGCCTTCTGGCCCTAAGGGCCTGGAGCCACCAGAAGAAGAGAAGAATCCCT 1523

Qy        1499 ACAAGGAAGTTTACACCGACATGTGGGTGGAGCCTGAAGCAGCTGCTTACGCCCCACCCC 1558
               |||| ||||| ||||| ||||||||||||||| ||||| ||||||||| ||||| ||||| |
Db        1524 ACAAAGAAGTGTACACGGACATGTGGGTGGAACCTGAGGCAGCTGCCTACGCACCACCTC 1583

Qy        1559 CACCAGCCAAGAAACCCAGAAAGAGCACAACAGAGAAACCTAAGGTCAAGGAGATCATTG 1618
               |||||||||| || ||| | ||||||||| | ||||| || ||||||||||||||| ||||
Db        1584 CACCAGCCAAAAAGCCCCGGAAGAGCACAGCGGAGAAGCCCAAGGTCAAGGAGATTATTG 1643

Qy        1619 ATGAGCGCACAAGGGAGCGGCTGGTGTATGAGGTGCGCCAGAAGTGCAGAAACATCGAGG 1678
               ||||||||||||| |||||||||||||||| ||||||||| ||||||||| | ||||| ||||
Db        1644 ATGAGCGCACAAGAGAGCGGCTGGTGTACGAGGTGCGGCAGAAGTGCCGGAACATTGAGG 1703

Qy        1679 ACATTTGTATCTCATGTGGGAGCCTCAATGTCACCCTGGAGCACCCACTCTTCATTGGAG 1738
               |||| || ||||| |||||||||||||||||||| ||||||||| ||||| ||||||| ||||||
Db        1704 ACATCTGCATCTCCTGTGGGAGCCTCAATGTTACCCTGGAACACCCCCTCTTCGTTGGAG 1763

Qy        1739 GCATGTGCCAGAACTGTAAGAACTGCTTCTTGGAGTGTGCTTACCAGTATGACGACGATG 1798
               | ||||||||| ||||| ||||||||||||  |||||||||| |||||||| |||||||| |
Db        1764 GAATGTGCCAAAACTGCAAGAACTGCTTTCTGGAGTGTGCGTACCAGTACGACGACGACG 1823

Qy        1799 GGTACCAGTCCTATTGCACCATCTGCTGTGGGGGGCGTGAAGTGCTCATGTGTGGGAACA 1858
               | |||||||||| |||||||||||||||||||| ||||| |||||||||||| || ||||
Db        1824 GCTACCAGTCCTACTGCACCATCTGCTGTGGGGGCCGTGAGGTGCTCATGTGCGGAAACA 1883

Qy        1859 ACAACTGCTGCAGGTGCTTTTGTGTCGAGTGTGTGGATCTCTTGGTGGGGCCAGGAGCTG 1918
               |||||||||||||||||||||| || |||||||||| ||||||||||||||| || ||||
Db        1884 ACAACTGCTGCAGGTGCTTTTGCGTGGAGTGTGTGGACCTCTTGGTGGGGCCGGGGGCTG 1943

Qy        1919 CTCAGGCAGCCATTAAGGAAGACCCCTGGAACTGCTACATGTGCGGGCATAAGGGCACCT 1978
               | ||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||
Db        1944 CCCAGGCAGCCATTAAGGAAGACCCCTGGAACTGCTACATGTGCGGGCACAAGGGTACCT 2003

Qy        1979 ATGGGCTGCTGCGAAGACGGGAAGACTGGCCTTCTCGACTCCAGATGTTCTTTGCCAATA 2038
               | ||||||||||||  | || || |||||||| || || |||||||||||||| || ||||
Db        2004 ACGGGCTGCTGCGGCGGCGAGAGGACTGGCCCTCCCGGCTCCAGATGTTCTTCGCTAATA 2063

Qy        2039 ACCATGACCAGGAATTTGACCCCCCAAAGGTTTACCCACCTGTGCCAGCTGAGAAGAGGA 2098
               |||| ||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||
Db        2064 ACCACGACCAGGAATTTGACCCTCCAAAGGTTTACCCACCTGTCCCAGCTGAGAAGAGGA 2123

Qy        2099 AGCCCATCCGCGTGCTGTCTCTCTTTGATGGGATTGCTACAGGGCTCCTGGTGCTGAAGG 2158
               |||||||||| ||||||||||||||||||||| || ||||||||||||||||||||||||||
Db        2124 AGCCCATCCGGGTGCTGTCTCTCTTTGATGGAATCGCTACAGGGCTCCTGGTGCTGAAGG 2183

Qy        2159 ACCTGGGCATCCAAGTGGACCGCTACATTGCCTCCGAGGTGTGTGAGGACTCCATCACGG 2218
               || ||||||| || |||||||||||||||||||| |||||||||||||||||||||||||
Db        2184 ACTTGGGCATTCAGGTGGACCGCTACATTGCCTCGGAGGTGTGTGAGGACTCCATCACGG 2243

Qy        2219 TGGGCATGGTGCGGCACCAGGGAAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACAC 2278
               |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        2244 TGGGCATGGTGCGGCACCAGGGGAAGATCATGTACGTCGGGGACGTCCGCAGCGTCACAC 2303

Qy        2279 AGAAGCATATCCAGGAGTGGGGCCCATTCGACCTGGTGATTGGAGGCAGTCCCTGCAATG 2338
               |||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||
Db        2304 AGAAGCATATCCAGGAGTGGGGCCCATTCGATCTGGTGATTGGGGGCAGTCCCTGCAATG 2363

Qy        2339 ACCTCTCCATTGTCAACCCTGCCCGCAAGGGACTTTATGAGGGTACTGGCCGCCTCTTCT 2398
               |||||||||| |||||||||||| ||||||||| || || |||||| |||||||| |||||||
Db        2364 ACCTCTCCATCGTCAACCCTGCTCGCAAGGGCCTCTACGAGGGCACTGGCCGGCTCTTCT 2423

Qy        2399 TTGAGTTCTACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCT 2458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2424 TTGAGTTCTACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAGATGATCGCCCCTTCT 2483

Qy        2459 TCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAGGGACATCTCGCGAT 2518
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2484 TCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAGGGACATCTCGCGAT 2543

Qy        2519 TTCTTGAGTCTAACCCCGTGATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGGCCC 2578
               |||| ||||| ||||| |||||||||||| |||||||||||||| ||||||||||||||||
Db        2544 TTCTCGAGTCCAACCCTGTGATGATTGATGCCAAAGAAGTGTCAGCTGCACACAGGGCCC 2603

Qy        2579 GTTACTTCTGGGGTAACCTTCCTGGCATGAACAGGCCTTTGGCATCCACTGTGAATGATA 2638
               | |||||||||||||||||||| || ||||||||||| ||||||||||||||||||||||
Db        2604 GCTACTTCTGGGGTAACCTTCCCGGTATGAACAGGCCGTTGGCATCCACTGTGAATGATA 2663

Qy        2639 AGCTGGAGCTGCAAGAGTGTCTGGAGCACGGCAGAATAGCCAAGTTCAGCAAAGTGAGGA 2698
               ||||||||||||| |||||||||||||||| ||||| ||||||||||||||||||||||||
Db        2664 AGCTGGAGCTGCAGGAGTGTCTGGAGCATGGCAGGATAGCCAAGTTCAGCAAAGTGAGGA 2723

Qy        2699 CCATTACCACCAGGTCAAACTCTATAAAGCAGGGCAAAGACCAGCATTTCCCCGTCTTCA 2758
               ||||||| || ||||||||||| ||||||||||||||||||||||||||||| || |||||||
Db        2724 CCATTACTACGAGGTCAAACTCCATAAAGCAGGGCAAAGACCAGCATTTTCCTGTCTTCA 2783

Qy        2759 TGAACGAGAAGGAGGACATCCTGTGGTGCACTGAAATGGAAAGGGTGTTTGGCTTCCCCG 2818
               |||| ||||| ||||||||| | ||||||||||||||||||||||| ||||| ||||| |
Db        2784 TGAATGAGAAAGAGGACATCTTATGGTGCACTGAAATGGAAAGGGTATTTGGTTTCCCAG 2843

Qy        2819 TCCACTACACAGACGTCTCCAACATGAGCCGCTTGGCGAGGCAGAGACTGCTGGGCCGAT 2878
               ||||||| || ||||||||||||||||||||||||||||||||||||||||||||||| |
Db        2844 TCCACTATACTGACGTCTCCAACATGAGCCGCTTGGCGAGGCAGAGACTGCTGGGCCGGT 2903

Qy        2879 CGTGGAGCGTGCCGGTCATCCGCCACCTCTTCGCTCCGCTGAAGGAATATTTTGCTTGTG 2938
               | |||||||||||| ||||||||||||||||||||||||||||||| ||||||||| ||||
Db        2904 CATGGAGCGTGCCAGTCATCCGCCACCTCTTCGCTCCGCTGAAGGAGTATTTTGCGTGTG 2963

Qy        2939 TGTAAGGGACATGGGGGCAAACTGAAGTAGTGA 2971
               ||||||||||||||||||||||||| |||| ||
Db        2964 TGTAAGGGACATGGGGGCAAACTGAGGTAGCGA 2996
```

```
RESULT 5
AK025230
LOCUS       AK025230                2191 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ21577 fis, clone COL06724, highly similar to
            AF067972 Homo sapiens DNA cytosine methyltransferase 3 alpha
            (DNMT3A) mRNA.
ACCESSION   AK025230
VERSION     AK025230.1  GI:10437699
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL06724.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (sites)
AUTHORS     Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 2191)
AUTHORS     Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE       Direct Submission
JOURNAL     Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..2191
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COL06724"
                     /tissue_type="colon"
                     /clone_lib="COL"
     misc_feature    1..2191
                     /note="cloning vector pME18SFL3"
                     /note="highly similar to AF067972 Homo sapiens DNA
                     cytosine methyltransferase 3 alpha (DNMT3A) mRNA."
BASE COUNT    571 a    505 c    575 g    540 t
ORIGIN
```

```
Query Match           37.0%;
Best Local Similarity  86.8%;
Matches 190; Conservative  0; Mismatches  224; Indels  65; Gaps  15;

              Score 1551.6;   DB  9;   Length 2191;
              Pred. No.0;
```

```
Qy  1615 ATTGATGAGCGCCACAAGGAGGACGGCGGCTCGGTCGTTGTATGAGGTGCGCCAAGTGCAGAAACATC 1674
         ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||
Db     2 ATTGATGAGCGCCACAAGGAGGACGGCGGCTCGGTGTGTACGAGGTGCGCCAGAAAGTGCGGAACATT 61

Qy  1675 GAGGACAATTTGTATCTCATGTGGGAGCGCCTCAAATGTCACCCGTCACAAATGGAATGGAAAAGTGT 1734
         |||||||||||| ||| ||||||||||||||| |||||| |||||||| ||||||||||| |||||||
Db    62 GAGGACAATCTGCATCTCTCGTGTGGGAGCGCCTCAAATGTTAACCCGTGAACACCCCTCTTCGTT 121

Qy  1735 GGAGGCCATGTGCCAGAACGTATAAGAAACTGCTTCTTCGGAGTGTGGCTTACCAGTATGACGAC 1794
         |||||| |||||||||||||||| |||||||||||||| |||||||||| |||||||||||||||
Db   122 GGAGGAATGTGCCAAAAACTGCAAGAAACTGCTTTCTGGAGTGTGGCGTACCAGTACGACGAC 181

Qy  1795 GATGGGTACCAGTCCTTATTGCACCATCTCTCGTGTGGGGGGGGGTGAAGTGGCTCAATGTTGTGGG 1854
         || | ||||||| |||||| ||||||||| |||||||||||||||||  | |||||||||| || |||
Db   182 GACGGCTACCAGTCCTACTGCACCATCTCGTGTGGGGGTGCTCATGTGCCGA 241

Qy  1855 AACAACAACTGCTGCAGGTGCTTTGTGTCGAGTGTGTGAGATCTCTGGGTGGGGGCCAGGA 1914
         |||||||||||||||||| |||||| |||| ||||||||||||| |||||||||||||||| ||
Db   242 AACAACAAACTGCTGCAGGTGCTTTTGCGTGGAGTGTGTGGAACCTCTTGGTGGGGGCCCGGG 301

Qy  1915 GCTGCTCAGGCACAGCCATTAAGGAAGAAGACCCCTGGGAACTGCTCTACATGTGCCGGCATAAGGGC 1974
         |||| ||||||||||||||||| ||||||||||||||||||||||||| ||| ||||||||||| |||
Db   302 GCTGCCCAGGGCAGCCATTAAGGAAGAAGACCCCTGGGAACTGCTACATGTGCCGGCACAAGGGT 361

Qy  1975 ACCTATGGGCTGCTGCGAAGACGGGAAGACTGGGCCCTTCTCCGACTCCAGATGTTCTTTTGCC 2034
         ||||||||||||||||| |||  |||||| ||||||||| |||| |||||||||||||||||||
Db   362 ACCTACGGGCTGCTGCGCGGCGGAGAGGACTGGCCCCTCCCGGACTCCAGAGTGTTCTTCAG 421

Qy  2035 AATAACCATGACCACGAGGAAATTTGACCCCCCAAAAGGTTTACCCCACCCTGTGCCACGTGAGAAG 2094
         |||||| |||||||| |||||||||||||||||  | |||||||||||| | ||||||||||||||||
Db   422 AATAACCACGACCAGGAATTTGACCCTCCAAAAGGTTTGACATCATCCTCCTGATTTTTCAATGTTATTCTTCAG 481

Qy  2095 AGGAAGCCCATCCGCGGTCGTGTCTCTTTGATGGGATTGCTACAGGCGCTCGTCGTGGTGCTG 2154
         ||||||||||||||||||||  |||||||||||||||||  | |||||||||||||| |||||
Db  3172 TCCTATCTAAAAGCAAAATAGGC-CCCTCCCCTTCTTCCCC---TCCGGTCCTAGGAGG 3227
```

```
Db   482 AGGAAGCCCATCCGGGTGCTGTCTCTCTTTGATGGAATCGCTACAGGGCTCCTGGTGCTG 541

Qy  2155 AAGGACCTGGGCATCCAAGTGACATGGAGCCGCTACATTGCTCCGAGCGTGTCTGAGGACTCCATC 2214
         ||||||  |||||||||| |||  |||||| |||||||||||||||||||||| || ||||||||
Db   542 AAGGACTTGGGCATTCAGGTGACCTGGAGCCGCTACATTGCCTCGGAGGTGTGTGAGGACTCCATC 601

Qy  2215 ACGGTGGGGCATGGTGCGGCGCCACCAGGGAAAGATCATGTACTCGGGGACGTCGGCAGCGTC 2274
         |||||||||||||| ||||| ||||||||||||||||||||||| ||||||||||| ||||||
Db   602 ACGGTGGGGCATGGTGCGGCACCAGGGAGAGATCATGTACGTCGGGGAGGACTCCATC 661

Qy  2275 ACACAGAAAGCACATATCCAAGGAGTGGGGCGGGCCCCATTCGAACCTGGTGGGGGGAGTGCCCTGC 2334
         |||| |||| |||||||| |||||||||||||||||||||||| | |||||| |||||||||||||
Db   662 ACACAGAAAGCATATCCAAGGAGTGGGGGCGGGGCCCCATTCGAACCTGGTGATTGGGGGGAAGTGCCCTGC 721

Qy  2335 AATGACCTCTCCATTGTCAACCCTGCCCCCAAGGGACTTTGATGAGGTGACTTGGCCGGCCTC 2394
         |||||||||||||| |||| |||||||||| ||||||||| |||||||| || ||||||||||
Db   722 AATGACCTCCATCGTCAACCCTGCCCCCAAGGGCCTCTACGAGGGCACTGGCCTC 781

Qy  2395 ACACAGAAAGCACATATCCAGGAGTGGGGGCCCCATTCGAATCTGGTGATTGGGGGGAAGTGCCCC 2454
         |||| |||| ||||||||| |||||||||||| ||||||| | ||||||||||||||||||||||
Db   782 ACACAGAAAGCATATCCAGGAGTGGGGGCCCCATTCGAATCTGGTGATTGGGGGGAAGTGCCCC 841

Qy  2455 TTCTCTCTGGCTCTTGAGAGTGTCGGTGCGCCATGGGCGTTAGTGACAAGGACATCTCG 2514
         |||||||||||||||||||||| ||||||| ||||||||||| |||||||||||||||
Db   842 TTCTCTCTGGCTCTTTGAGAAGTGTCTGGAGCATGGCAGGGATAGCCAAGTTCAGCAAAGTG 901

Qy  2515 CGATTTCTTGAGTCTAACCCGTGATGATTGACGCGCACTGAAATGGAAAGGAGAGTGCCCTTC 2574
         ||||||  |||||||||||||| |||| |||| ||||||||||||||||||||||||||||||
Db   902 CGATTTCTCGAGTCCAACCCTGTGATGATTGACTGAAATGGAAAGAGAGTGTTTTGGGCTTC 961

Qy  2575 GCCCGTTACTTCTCTGGGGTAAACTTCCTGCCATGAACAGGCCTTTTGGCATCCACTGTGAAT 2634
         ||||||||||||| ||||| |||||||||| |||||||| |||||||| ||||||||||
Db   962 GCCCGCTACTTCTCTGGGTAAACCTTCCCGTGAACAGGCCGTTGCATCCACTGTGAAT 1021

Qy  2635 GATAAGCTGGAGCTGCAAGAGTGTCTGGACGACGGCAGAATAGCCAAGTTCAGCAAAGTG 2694
         ||||||| ||||||||||||||| ||||| || |||||||||||||||||||||||||||
Db  1022 GATAAGCTGGAGCTGCAGGAGTGTCTCGAGGAGCATGGCAGAGGATAGCCAAGTTCAGCAAAGTG 1081

Qy  2695 AGGAACCATTACCACCAGTCAAACTCTATAAAAGCAGAGAACCAGAGCATTCCCCGTC 2754
         |||||||| | ||||||||||||||||||||||||||||||||| || ||||||||||
Db  1082 AGGACCATTACTACGAGGTCAAACTCCATAAAAGCAGAGAACCAGCATTTTCCTCGTC 1141

Qy  2755 TTCATGAAACGAGAAGAGAGACATCCTGTGGTGCACTGAAATGGAAAAGGAGGTGTTTGGGCTTC 2814
         ||||||| | ||||||||||| |||||||||||||||||||| ||||||||||||||||
Db  1142 TTCATGAAATGAGAAGAGAGGACACATCTTATGGTGCGCACTGAAATGGAAAAGGGTATTGGGTTTC 1201

Qy  2815 CCCGTCCACTACACAGACGTCTCCAACATGAGCCCCTTGGCGGAGGCAGAGACTGCTGGGC 2874
         || ||||||| || ||||| |||||| ||||||||||| ||||||||||||||||||||||
Db  1202 CCAGTCCACTATACTGACGTCTCCAACATGAGCCCGCTTGGCGGAGGCAGAGACTGCTGGGC 1261

Qy  2875 CGATCGTGGAGGCGGTGCCGTGTCATCCGCGCCACTCTTCGCTCCGCTGAAGGAATATTTGCT 2934
         |||| ||||||||||||||||||||||||||||||||||||  |||||| | |||||||||||
Db  1262 CGGTCATGGAGCGGTGCCCAGTCATCGCGCCAACTCTTCGCTCGCTGAAGGAGTATTTTGCG 1321

Qy  2935 TGTGTGTAAGGGACATGGGGGGCAAAACTGAAGTAGTGATGATAAAAAAGTTAAACAAAACA 2994
         ||||| | |||||||||||||||||||||||| |||||| | |||||||| |||| |||||
Db  1322 TGTGTGTTAAGGGACATGGGGCAAAACTGAGGTAGGGA----CACAGAAGTTAAACAAACAA 1377

Qy  2995 ACAAACAAAAAAACAAAACAAAACAATAAAACACCAAGACAGAGCACGGAGAAAAGT-T 3053
         ||| |||| || ||||||||||||| ||||||||||||||||||||||||||||  |
Db  1378 AC---AC-----AAAAACACAAAAACATAATAAAACACCAAGAACATGGAGGAAGAAGTAT 1432

Qy  3054 CAGCACCCAGAGAGAGAGAGAGAAAAAGGAATTTAAAGC-- -AAACCACAGAGAGGAAAAACGCCGGA 3111
         ||||||||||||||||||||||||||||||||||||| ||
Db  1433 CAGCACCCAGAGAGAGAGAGAGAGATTTAAAACACAGAGCGGAAAAATAACCGGAG 1492

Qy  3112 GGGCTTGCGCCTTGCAAAAGGGTTTGGACATCATCATCTCCTGAGTTTTCAATGTTATCTTCAG 3171
         ||||||||||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||
Db  1493 GGCTTTGCCTTGCGAAAAGGGTTGGACATCAATCATCTCCTGATTTTTCAATGTTATTCTTCAG 1552
```

RESULT 6

```
LOCUS       AC118195
DEFINITION  AC118195              204534 bp    DNA     linear    HTG 17-APR-2002
            Mus musculus clone RP23-9D1, WORKING DRAFT SEQUENCE, 11 ordered
            pieces.
ACCESSION   AC118195
VERSION     AC118195.2  GI:20163100
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 204534)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204534)
```
```
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
          Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
          Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
          Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
          Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
          Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,L.,
          McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
          Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
```
```
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Nell,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2002) Whitehead Institute/ MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 204534)
```
```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,L.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Nell,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-2002) Whitehead Institute/ MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 17, 2002 this sequence version replaced gi:20147966.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
              Center project name: L19641
              Center clone name: 9_D_1
            --------------- Summary Statistics
              Sequencing vector: plasmid; n/a; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 201846 bases at least Q40
              Consensus quality: 202797 bases at least Q30
              Consensus quality: 203171 bases at least Q20
              Insert size: 210000; agarose-fp
              Insert size: 203534; sum-of-contigs
              Quality coverage: 8.0 in Q20 bases; agarose-fp
              Quality coverage: 8.3 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *      1      9535: contig of 9535 bp in length
            *   9536   9635: gap of            100 bp
```

Sequence alignment region:

```
Db  1553  TCCTATTTAAAACAAAACCAAGCTCCCTTCCCTTCCCTCCCCCCTTCCCTTTCCTTTTTTTTTTTCGGT 1612
Qy        |||||| |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3228  CGAACATTTTTGTTTTCTACTCTCTTTTTCTACAGGGCGTTTTCTGTTTTGTTTTTGGGTTTTTTTGTTT 3287
Db        |||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||||
Db  1613  CAGACCTTTTATTTTTCTTACTC-TTTTCAGAGGGGTTTTCTGTTTCGTTTGTTTGTTTGGG-TTTTCGTTT 1670
Qy        ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Qy  3288  CTTGCTGTGACTGAAAACAAGAGAGAGTTATTGCCAGC-AAAATCAGTAACAACAAAAGTAGA 3346
Db        |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||| |||
Db  1671  CTTGCTGTGACTGACAAAAGGTTATTGCCAGCAAAAATCAGTAACAAAAAATAGTAAC 1730
Qy  3347  AATGCCTTGGAGAGAAAGG-----------GAGAGAGGGAAAAATTCTATAAAAAACTTAAAA 3397
Db        ||| |||| |||||||||||         |||||||||||||||||||||||||| |
Db  1731  AATACCTTGCAGAGGGAAAAGTGGGGAGAGAGGAAAAAAGGAAAATTCTATAGAATCTATATA 1790
Qy  3398  TATTGGTTTTTTTTTTTTTTTTTTCCTTT---------------------------------TCTA 3426
Db        | ||||||||||||||||||||| |||                              | ||
Db  1791  TTGGGTTGTTTTGTTTTTTTTGTTTTTTGTTTTTTGGGTTTTTTTTTTTTTTTTTTTACTATA 1850
Qy  3427  TATATCTCTTTTGGGTGTGTCTCTTAGCCCTGATCAGATAGGAGGCACAAACA--GGAAGAGAATA 3484
Db        |||||||||||||||||||||| |||||||||||||||||||||||||||||  ||||| |||
Db  1851  TATCTTTTTTTTTGTTGTTGTCTCTRAGCCTGATCAGATAGGAGCACACAAGCACAGGGGACGGGAAAGA 1910
Qy  3485  GAGACCCTCGGAGGCAG--AGTCTCCTCTCCCCACCCCCCGAGCAGTCTCAACACGCACCATT 3543
Db        ||| |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db  1911  GAGAGACAGCTCAGGCGGCGCAGCAGCATTCCCTCCCCAGCCCACTGAGCCTGTCGTGCCGCACCAATT 1970
Qy  3544  CCTGGTCATGCAAAAACAGAACCCAACTAGCAGCAGGCGGCCTGAGAGAACACCACACCAGA 3603
Db        ||||||||| |||||||||||||||||| |||||||||||||||||||||||||||||||| |
Db  1971  CCTGGTCACGGAAAAACAGAAACCCAGTTAGCAGCAGGGAGCAGGAGAAAACACCACACAAGACA 2030
Qy  3604  CACTTTCTACAGTATTTCAGGTGCCTACCACACAGGAAACCTTGANAGAAAAACCAGTTTCT 3663
Db        |||||||||||||||||||||||||| ||||||||||||||||||| |||||||||||||||
Db  2031  -TTTTCTCACAGTATTTCAGGTGCCTACCACACAGGAAAACCTTGAAAGAAAATCAGTTTCT 2089
Qy  3664  AGAAGCCCCTGTTACCTCTTGTTGTTTACAGTTTATATATATATATARGAATAGAATATATGAGATATATATA 3723
Db        ||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db  2090  AGAAGCCCCTGTTACCTCTTGTTTACAGTTTTATATATAGATGATAGATAGTAG--ATATA 2147
Qy  3724  TATATAAAAAGGTACTGTTAACTACTGTACA 3753
Db        |||||||||||||||||||||||||||||||
Db  2148  TATATAAAAGGTACTGTTAACTACTGTAAA 2177
```

```
FEATURES             Location/Qualifiers
    source           1..204534
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-9D1"
                     /clone_lib="RPCI-23 Female Mouse BAC"
    misc_feature     1..9535
                     /note="assembly fragment
                     clone_end:SP6
                     vector_side:left"
    misc_feature     9636..11346
                     /note="assembly fragment"
    misc_feature     11447..13984
                     /note="assembly_fragment"
    misc_feature     14085..17045
                     /note="assembly_fragment"
    misc_feature     17146..34039
                     /note="assembly_fragment"
    misc_feature     34140..49222
                     /note="assembly_fragment"
    misc_feature     49323..67306
                     /note="assembly_fragment"
    misc_feature     67407..110848
                     /note="assembly_fragment"
    misc_feature     110949..147393
                     /note="assembly_fragment"
    misc_feature     147494..180251
                     /note="assembly_fragment"
    misc_feature     180352..204534
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"

*    9636  11346: contig of 1711 bp -in length
*   11347  11446: gap of        100 bp
*   11447  13984: contig of 2538 bp in length
*   13985  14084: gap of        100 bp
*   14085  17045: contig of 2961 bp in length
*   17046  17145: gap of        100 bp
*   17146  34039: contig of 16894 bp in length
*   34040  34139: gap of       100 bp
*   34140  49222: contig of 15083 bp in length
*   49223  49322: gap of       100 bp
*   49323  67306: contig of 17984 bp in length
*   67307  67406: gap of       100 bp
*   67407 110848: contig of 43442 bp in length
*  110849 110948: gap of       100 bp
*  110949 147393: contig of 36445 bp in length
*  147394 147493: gap of       100 bp
*  147494 180251: contig of 32758 bp in length
*  180252 180351: gap of       100 bp
*  180352 204534: contig of 24183 bp in length.
```

```
BASE COUNT    52148 a   49228 c   50043 g   52097 t   1018 others
ORIGIN
```

```
Query Match      32.7%;  Score 1369.4;  DB 2;  Length 204534;
Best Local Similarity  99.7%;  Pred. No. 0;
Matches 1392; Conservative  0;  Mismatches  2;  Indels  2;  Gaps  2;
```

Qy   4117   CAGAAAATATTCCAGAGTGAGGGAAAAGTGACCCGCGCCATTAACCCCNCCTGGAGCAAATAA 4176
    |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   164017   CAGAAAATATTCCAGAGTGAGGGAAAAGTGACCCGCGCCATTAACCCCACCTGGAGCAAATAAA 164076

Qy   4177   AAAACATACAAAATGT 4192
    ||||||||||||||||
Db   164077   AAAACATACAAAATGT 164092

RESULT 7
AC111092
LOCUS   AC111092   208874 bp   DNA   linear   HTG 24-MAY-2002
DEFINITION   Mus musculus clone RP23-16809, WORKING DRAFT SEQUENCE, 9 ordered
     pieces.
ACCESSION   AC111092
VERSION   AC111092.3   GI:21166254
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Mus musculus
  ORGANISM   Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1   (bases 1 to 208874)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Direct Submission
  JOURNAL   Mus musculus, clone RP23-16809
    Unpublished
REFERENCE   2   (bases 1 to 208874)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
    Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
    Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
    Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
    Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
    Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
    Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
    Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
    Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
    Landers,T., Lehoczky,J., Levine,R., Liu,G., Maclean,C.,
    Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
    McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
    Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
    Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
    Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
    Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
    Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
    Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
    Topham,K., Travers,M., Travis,N., Trigillo,J., Vassiliev,H.,
    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
    Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3   (bases 1 to 208874)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
    Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
    Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
    Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
    Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
    Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
    Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
    Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
    Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
    Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
    Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
    Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
    Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
    Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
    O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
    Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
    Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
    Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
    Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
    Theodore,J., Topham,K., Travers,M., Travis,N., Trigillo,J.,
    Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Qy   4117   CAGAAAATATTCCAGAGTGAGGGAAAAGTGACCCGCGCCATTAACCCCNCCTGGAGCAAATAA 4176

Db   164017   CAGAAAATATTCCAGAGTGAGGGAAAAGTGACCCGCGCCATTAACCCCACCTGGAGCAAATAAA 164076

Qy   4177   AAAACATACAAAATGT 4192

Db   164077   AAAACATACAAAATGT 164092

    Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 24, 2002 this sequence version replaced gi:21104982.
    All repeats were identified using RepeatMasker:
    Smit, A.F.A. & Green, P. (1996-1997)
    http://ftp.genome.washington.edu/RM/RepeatMasker.html
    ------------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
    ------------------- Project Information
    Center project name: L22454
    Center clone name: 168_O_9
    ------------------- Summary Statistics
    Sequencing vector: Plasmid; n/a; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 206101 bases at least Q40
    Consensus quality: 207369 bases at least Q30
    Consensus quality: 207812 bases at least Q20
    Insert size: 210000; agarose-fp
    Insert size: 208074; sum-of-contigs
    Quality coverage: 8.8 in Q20 bases; agarose-fp
    Quality coverage: 8.9 in Q20 bases; sum-of-contigs
    ----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 9 contigs. Gaps between the contigs
    * are represented as runs of N. The order of the pieces
    * is believed to be correct as given, however the sizes
    * of the gaps between them are based on estimates that have
    * provided by the submitter.
    * This sequence will be replaced
    * by the finished sequence as soon as it is available and
    * the accession number will be preserved.
    *   1   2217: contig of 2217 bp in length
    *   2218   2317: gap of   100 bp
    *   2318   4169: contig of 1852 bp in length
    *   4170   4269: gap of   100 bp
    *   4270   5258: contig of 989 bp in length
    *   5259   5358: gap of   100 bp
    *   5359   16517: contig of 11159 bp in length
    *   16518   16617: gap of   100 bp
    *   16618   27463: contig of 10846 bp in length
    *   27464   27563: gap of   100 bp
    *   27564   53595: contig of 26032 bp in length
    *   53596   53695: gap of   100 bp
    *   53696   80829: contig of 27134 bp in length
    *   80830   80929: gap of   100 bp
    *   80930   170253: contig of 89324 bp in length
    *   170254   170353: gap of   100 bp
    *   170354   208874: contig of 38521 bp in length.
FEATURES   Location/Qualifiers
  source   1..208874
          /organism="Mus musculus"
          /db_xref="taxon:10090"
          /clone="RP23-16809"
          /clone_lib="RPCI-23 Female Mouse BAC"
  misc_feature   1..2217
          /note="assembly_fragment
          clone_end:SP6
          vector_side:left"
  misc_feature   2318..4169
          /note="assembly_fragment"
  misc_feature   4270..5258
          /note="assembly_fragment"
  misc_feature   5359..16517
          /note="assembly_fragment"
  misc_feature   16618..27463
          /note="assembly_fragment"
  misc_feature   27564..53595

```
misc_feature    53696..80829
                /note="assembly_fragment"
misc_feature    80930..170253
                /note="assembly_fragment"
misc_feature    170354..208874
                /note="assembly_fragment"
                /note="assembly_fragment"
                clone_end:T7
                vector_side:right"

BASE COUNT    53835 a    50502 c    49598 g    54139 t
ORIGIN                                                      800 others

Query Match          32.7%;  Score 1369.4;  DB 2;  Length 208874;
Best Local Similarity    99.7%;  Pred. No. 0;
Matches 1392;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy    2798 AAAGGGTGTTTGGCTTCCCCCGTCCCACTCACACACGTCTCCAACATGAGCGCGTTGGGGA 2857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94876 ACAGGGTGTTTGGCTTCCCCCGTCCCACTCACACACGTCTCCAACATGAGCGCGTTGGGGA 94935

Qy    2858 GGCAGAGACTGCTGGGCGCCGATCGTGTGGAGGTGCGCGGTCATCCGCCACCTCTTCGCTCCGC 2917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94936 GGCAGAGACTGCTGGGCGCCGATCGTGTGGAGGTGCGCGGTCATCCGCCACCTCTTCGCTCCGC 94995

Qy    2918 TGAAGGAATATATTTTGCTTGTGTGTAAGGGACANATGGGGACTGAAGTAGTAGTGATGATGA 2977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95056 TGAAGGAATATATTTTGCTTGTGTGTAAGGGACANATGGGGACTGAAGTAGTAGTGATGATGA 95055

Qy    2978 AAAAGTTAAACAAACAAACAAACAAACAAACAATAAAACACCAAGAACGAG 3037
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95056 AAAAGTTAAACAAACAAACAAACAAACAAACAATAAAACACCAAGAACGAG 95115

Qy    3038 AGGACGGAGAAAAGTTCAGCACCCCAGAAGAGAGAAAAAGGAAATTAAAAGTGATGATGA 3097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96075 AGGACGGAGAAAAGTTCAGCACCCCAGAAGAGAGAAAAAGGAAATTAAAAGTGATGATGA 96134

Qy    3098 AGGAAAAACGCCGGAGGGGCTTGGGCCTTGCAAAAAGGGGT 3157
           |||||||||||||||||||||||||||||||||||||||||
Db   95176 AGGAAAAACGCCGGAGGGGCTTGGGCCTTGCAAAAAGGGGT 95235

Qy    3158 ATGTTAACCTTCAGTCCTATCTATAAAAACGCAAAAATAGGCCCCTCCCCCT 3217
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95236 ATGTTAACCTTCAGTCTCTATCTATAAAAACGCAAAAATAGGCCCCTCCCCCT 95295

Qy    3218 TCCTAGGAGGCGGCGAACTTTTTGTTTTCTACTCTTTTTCAGAGGGT 3277
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   95296 TCCTAGGAGGCGGCGAACTTTTTGTTTTCTACTCTTTTTCAGAGGGT 95355

Qy    3278 GTTTTGTTTTCTTGCGTGTGACTGAAACAAGAGAGTGTTATTGCCAGCAAAATCAGTAACAACA 3337
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95356 GTTTTGTTTTGTTCTGCTGTGACTGAAACAAGAGAGTTATTGCAGCAAAATCAGTAACAACA 95415

Qy    3338 AAAAGTAGAAAATGCCTTGGAGAGGAGGAGGAGGAGGAAAAATTCTATAAAAAACTTAAAA 3397
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95416 AAAAGTAGAAAATGCCTTGGAGAGGAGGAGGAGGAAAAATTCTATAAAAAACTTAAAA 95475

Qy    3458 GATAGGAGCACAAAACAGGAAGAGAAGTAGAGACACCTCGGACGTCTCCTCTCCCACC 3517
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95476 TATTGG-TTTTTTTTTTTTTTTTCCTTTTCTATATATATCTCTTTCAGGTGGTCTGTGTCTCTCAGGCTGATCA 95534

Qy    3518 CCCCGAGCAGTCTCAACAGCACCACCAGACAC-TTTCTACAGTATATTTCAGGTGCTACCACAC 3577
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   95535 GATAGGAGCACAAAACAGGAAGAGAAATAGGAGACCCTCGGAGCGAGTGTCTTCCTCTCCCACC 95594

Qy    3578 GGGGCCCTGAGAGAACACCACCAGAACAC-TTCTACAGTATTTCTACAGTGTGCCTACCCACAC 3636
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   95595 CCCCGAGCAGTCGTCAACAGCACCATTCCTGGTCATGCAAAACAGCCAACTAGCACGA 95654

Qy    3637 AGGAAAACCTTGAAGAAAACCAGTTTCTAGAAGCCGCCTGTTACCCTCTGTTTACCACAGTTTAT 3696
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95655 GGGGCCGGCTGAGAGAAACACCACCACACACTTCCTACAGTATTTCAGGTGCCTACCCACAC 95714
```

```
Db   95715 AGGAAAACCTTGAAGAAAACCAGTTTCTAGAAGCCGCCTGTTACCCTCTGTTTACCACAGTTTAT 95774
Qy    3697 ATATATATGATAGATATTGAGATATATATATATATAAAAGGTAGTGTTAACTACTGTCATTCC 3756
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95775 ATATATATGATAGATATGAGATATATATATATAAAGGTAGTGTTAACTACTGTCANCC 95834
Qy    3757 CGACTTCATAATGGTGCTTTCAAAACACGAGATGAGAGATGACAAGACATCAGCTTCCGCCTGG 3816
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95835 CGACTTCATAATGGTGCTTTCAAAACACGAGATGACAAGACATCAGCTTCCGCCTGG 95894
Qy    3817 CCCTCTGTTGCAAAGGGTTTCAGCCCCAGGATGGGGGAGGAGGAGCAGCTGGAGGGGGTTTT 3876
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95895 CCCTCTGTTGCAAAGGGTTTCAGCCCCAGGATGGGGGAGGAGGAGCAGCTGGAGGGGGTTTT 95954
Qy    3877 AACAAAACTGAAGGAGGATGACCCATTATCACCCCCCCCCGCCCATGCCTAGTTCACCTG 3936
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95955 AACAAAACTGAAGGAGGATGACCCATCACCCCCCCCCGCCCATGCCTAGTTCACCTG 96014
Qy    3937 CCAAAAAGGGGGCTCAGCGTGAGGTGGTCGGCGGGGAAGCTGAGTGTGGAATTTATCC 3996
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96015 CCAAAAAGGGGGCTCAGCGTGAGGTGGTCGGCGGGGAAGCTGAGTGTGGAATTTATCC 96074
Qy    3997 AGACTCGCGTGCAATAACCTTAGNATNATGANATNTCNNANNTGACTGCCTTCAGAAAAATGGCT 4056
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96075 AGACTCGCGTGCAATAACCTTAGNATNATGAATCTNANATGACTGCCTTCAGAAAAATGGCT 96134
Qy    4057 TGAGAAAAACANTGTCCCCTGANTTTGAANTNCGTNCAGCCACGTTGANGGCCCCCCTTCCGGAT 4116
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96135 TGAGAAAAACANTGTCCCCTGATTTTGAANTNCGTNCAGCCACGTTGANGGCCCCCCTTCCGGAT 96194
Qy    4117 CAGAAATATTCCAGAGTGAGGGAAAGTGACCCGGCCCCATTAACCCNCCTGGAGCGAAATAAA 4176
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96195 CAGAAATATTCCAGAGTGAGGGAAAGTGACCCGGCCCCATTAACCCCACCTGGAGCAAAATAAA 96254
Qy    4177 AAAACATAACAAAAATGT 4192
           ||||||||||||||||||
Db   96255 AAAACATAACAAAAATGT 96270
```

RESULT 8
LOCUS   AC120824   129936 bp   DNA   linear   HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-282D16, *** SEQUENCING IN PROGRESS
            ***, 38 unordered pieces.
ACCESSION   AC120824
VERSION   AC120824.2 GI:21902541
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1 (bases 1 to 123936)
AUTHORS   Mizny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,R.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,K.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

```
TITLE
JOURNAL
REFERENCE
AUTHORS
```

            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.;
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.;
            Li,J., Li,Z., Lichtarge,O., Liou,C., Liu,J., Liu,W., Loulseged,H.;
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.;
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.;
            Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.;
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.;
            Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.;
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.;
            Oraqunye,N., Oviedo,R., Pace,A., Paycon,B., Peery,J., Perez,L.;
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.;
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.;
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.;
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.;
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.;
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.;
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.;
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.;
            Williams,G., Williamson,A., Wleczyk,R., Wooden,C., Worley,K.;
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.;
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 123936)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 123936)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Jul 18, 2002 this sequence version replaced gi:20531783.
COMMENT   -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center project name: GWLV
            Center clone name: CH230-282D16
            -------------- Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly Program: Phrap; version 0.990329
            Consensus quality: 95907 bases at least Q40
            Consensus quality: 99223 bases at least Q30
            Consensus quality: 101036 bases at least Q20
            ----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1   1009: contig of 1009 bp in length
            *   1010   1109: gap of unknown length
            *   1110   2113: contig of 1004 bp in length
            *   2114   2213: gap of unknown length
            *   2214   3233: contig of 1020 bp in length
            *   3234   3333: gap of unknown length
            *   3334   4442: contig of 1109 bp in length
            *   4443   4542: gap of unknown length
            *   4543   6202: contig of 1660 bp in length
            *   6203   6302: gap of unknown length
            *   6303   7464: contig of 1162 bp in length.
            *   7465   7564: gap of unknown length
            *   7565   8772: contig of 1208 bp in length
            *   8773   8872: gap of unknown length
            *   8873 10371: contig of 1499 bp in length
            * 10372 10471: gap of unknown length
            * 10472 11714: contig of 1243 bp in length
            * 11715 11814: gap of unknown length
            * 11815 12924: contig of 1110 bp in length
            * 12925 13024: gap of unknown length
            * 13025 14347: contig of 1323 bp in length
            * 14348 14447: gap of unknown length
            * 14448 15652: contig of 1205 bp in length
            * 15653 15752: gap of unknown length
            * 15753 17796: contig of 2044 bp in length
            * 17797 17896: gap of unknown length
            * 17897 18937: contig of 1041 bp in length
            * 18938 19037: gap of unknown length
            * 19038 21543: contig of 2506 bp in length
            * 21544 21643: gap of unknown length
            * 21644 24540: contig of 2897 bp in length
            * 24541 24640: gap of unknown length
            * 24641 28112: contig of 3472 bp in length
            * 28113 28212: gap of unknown length
            * 28213 31101: contig of 2889 bp in length
            * 31102 31201: gap of unknown length
            * 31202 34415: contig of 3214 bp in length
            * 34416 34515: gap of unknown length
            * 34516 37353: contig of 2838 bp in length
            * 37354 37453: gap of unknown length
            * 37454 40938: contig of 3485 bp in length
            * 40939 41038: gap of unknown length
            * 41039 44331: contig of 3293 bp in length
            * 44332 44431: gap of unknown length
            * 44432 48035: contig of 3604 bp in length
            * 48036 48135: gap of unknown length
            * 48136 51181: contig of 3046 bp in length
            * 51182 51281: gap of unknown length
            * 51282 55361: contig of 4080 bp in length
            * 55362 55461: gap of unknown length
            * 55462 59580: contig of 4119 bp in length
            * 59581 59680: gap of unknown length
            * 59681 63525: contig of 3845 bp in length
            * 63526 63625: gap of unknown length
            * 63626 67434: contig of 3809 bp in length
            * 67435 67534: gap of unknown length
            * 67535 72194: contig of 4660 bp in length
            * 72195 72294: gap of unknown length
            * 72295 77677: contig of 5383 bp in length
            * 77678 77777: gap of unknown length
            * 77778 81349: contig of 3572 bp in length
            * 81350 81449: gap of unknown length
            * 81450 86396: contig of 4947 bp in length
            * 86397 86496: gap of unknown length
            * 86497 91224: contig of 4728 bp in length
            * 91225 91324: gap of unknown length
            * 91325 97769: contig of 6445 bp in length
            * 97770 97869: gap of unknown length
            * 97870 105670: contig of 7801 bp in length
            * 105671 105770: gap of unknown length
            * 105771 111368: contig of 5598 bp in length .
            * 111369 111468: gap of unknown length
            * 111469 116680: contig of 5212 bp in length
            * 116681 116780: gap of unknown length
            * 116781 123936: contig of 7156 bp in length.
FEATURES   Location/Qualifiers
     source   1..123936
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-282D16"
BASE COUNT   35904 a 21413 c 21468 g 36486 t 8665 others
ORIGIN

Query Match   31.8%; Score 1331.8; DB 2; Length 123936;
```

Best Local Similarity  85.68%;  Conservative   0;  Mismatches  217;  Indels  33;  Gaps  5;

Matches 1485;  Conservative  0;  Pred No. 9.2e-307;  Mismatches 217;  Gaps  5;

```
Qy    843  GGCTGAGAAGAAAGCAAAGGTAATTGCACTAATGAAATGCTGTGGAAGAGAACCAGGCCTC  902
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84794  GGTTGAGAAGAAAGCCAAAGGTAATTGATTGCAGTAATCAAATGCTGTGGAGGAAACCAGGCCTC  84853

Qy    903  TGGGAGGTCTCAGAAAGTGGGAGGTGGGAGGCGCCAGCCCTTCCTCGTCTGTGCCAGCCCACCGACC  962
           ||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  84854  TGGGGAGTCTCAGAAAGGTGGAGCGTGGGAGGCGGCCAGCCCGTCCTCAGTGCAGCAGCCCA--GGACC  84912

Qy    963  TGCTTCTCCGACTGTGGCCACACCCCTGAGCCAGTAGGAGGGGATGCTGGGGACAAGAA  1022
           ||  ||||||||||| | ||||||||||||||||||||||||| ||||||||||||||||
Db  84913  TGCATCTCCACTGTGGCCCATTACCCCTGTGCCAGTAGGGGGCTCGATGCTGGGGGACAAGAA  84972

Qy   1023  TGCTACCAAAGCAGCCGACGATCAGCCTCAGTATGAGGGATGGCCGGGGGCTTTGGCATTGG  1082
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  84973  TGCTATCAAGGCAGCTGATGAGCAGCACCCCCAGTATGAGGAATGGCCTTGGGCATTGG  85032

Qy   1083  AGAGCTGGTGGTGTGGGGGAAACTTCGGGGGCTTCTCCTGGGCCAGGCCGAATTGTGTCTTG  1142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85033  AGAGCTGGTGTGGGGGAAACTCCGGGGCTTCTCCTGGGTGGGCCAAGCCGAAATTGTGTCTTG  85092

Qy   1143  GTGGATGACAGGCCCGGAGCCCGAGCCACTCGCTCGTCTGGGTCAGTGTGGTTCGGAGA  1202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85093  GTGGATGACAGGCCAGACCAAGGCCAGAGCCAGTCGCTCTGCATGTGTTTGGACA  85152

Qy   1203  TGGCAAGTTCTCAGTGGTGGTGTGTGTGGAGAAGCTCATGCCGCGTGAGCTGTCTTCTGCAGTGC  1262
           ||||||||||| | || |||||||||||||||||||||||||||||||||||||||||||||
Db  85153  TGGCAAGTTTTCAGTGATGTGTGTCGAGAAGCTCATGCCCGTCCTTCTCTGCA----  85208

Qy   1263  ATTCCACCAGGCCACTTACAACAAGCAGCCCCATGTACCGGCAAAGCCATCTACGAAGTCCT  1322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85209  TTTCCACCAGGCCACCTACAACAAGCAGCCCCATGTACCGCAAAAGCCATCTACGAAGTCCT  85268

Qy   1323  CCAGGTGGCCAGCACGCCGTGCCCGGGGAAGCTGTGTTCCAGCTTGCCATGACAGTG--ATGAA  1380
           ||||||| | ||||||||||||||||||||||||||||||||| ||||||     ||||
Db  85269  CCAGGTGGCCCAGCAGCTCGTTCAGGGGAAGCTGTTCCCAGCGCGTGTCATGAACAGCGACATGAA  85328

Qy   1381  AGTGACAGTGGCAAGGCTGTGTGGAAGTGCAGAACAACAGCAGATGATTGAATGGGCCTCGGT  1440
           ||||||||||||||||||||||||||||||||||||||| || |||||||||||||||||||
Db  85329  AGTGACACTGGGCAAGGCTGTGGAGGTGCAGAACAACAAGCACGAATTGAGTGGGTCCTTGGC  85388

Qy   1441  GGCTTCCAGCCCTCGGGTCCTTAAGGGCCTGGAGCCCACCCAGAGAGAAGAAGAAATCCTTAC  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  85389  GGCTTCCAGCCCTCTGGTTCCCAAGGGCCTGGAGCCCACCAGAGAGAAGAAGAAATCTTTAC  85448

Qy   1501  AAGGAAGTTTACACCGGACACATGTGGGTGGGGCGACAGTGTCACCCTGGAGCACCCACCCCCCA  1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85449  AAGGAAGTTTACACCGGACACATGTGGGTGGGGCGCTGCTGCAGCCTGCTATTGTCCCACCCCCA  85508

Qy   1561  CCAGCCAAGAAACCCAGAAAGACCTAAGGTCAAGGAGAGATCATTGAT  1620
           ||||||||||||||||||||||||||||||||||||||||||
Db  85509  CCAGCCAAGAAAACCCAGAAAGAGCAC------AACCTAAGGTCAAGGAGGATCATTGAT  85560

Qy   1621  GAGGCGCACAAGGGAGGACGGCGGGCTGGTTGGTGTATGAGGGTGCGCGCCCAGAGAAGTGCAGAAACATCGAGGAC  1680
           || | ||||||||||||| || |||||||||||| |||||||||||
Db  85561  GAACACACAAGGAACGACTCAATCATGCTAATGAGGTGCGCCAGAGAAGTGCCAAAACATCGAGGAC  85620

Qy   1681  ATTTTGTATCTCATGTGGGAGGCCTCAATGTCACCCTGGAGCACCCACCACTTGGGAGGC  1740
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |
Db  85621  ATTTTGTATCTCATGTGGGAGCCTCAATGTCAACCTTGGAGCACCACTCGTTGGGAGGA  85680

Qy   1741  ATGTGCCAGAACTGTAAGAAACTGCTTCTTGGAGTGTGTGCTTACCAGTATGACGGATGGG  1800
           ||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db  85681  ATGTGCCAGAACTGTAAGAAACTCGCTTCTCTGGAGTGTGTCTTACCAATACGATGACGGATGTG  85740

Qy   1801  TACCAGTCCTATTGCACCATCTGCTGTGTGGGGGAACAAC  1860
           || ||| || | |||||||||||||||||||||||||||||
Db  85741  TACCAGTCCTTCTGTGTACCAATCTGCTGTTTGGGGGAACAAC  85800

Qy   1861  AACTGCTGCAGGTGCTTTTGTGTCGGAGTCTCTTGGGTGTGGGGCCAGGGAGCTGCT  1920
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db  85801  AACCGGCTGCAGGTGCTGCTTTGTGTGGGGAGCTGTGGCTGGGATCTCCTTGGGTGTGGGGCCAGGGCACCTNT  85860
```

```
Db  85801  AACCGGCTGCAGGTGCTGCTTTGTGTGGGGGAGCTGTGGGGGATCTCTTGGGTGTGGGGCCAGGGCACCTNT  85860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1921  CAGGCAGCCATTAAGGAAGACCCCTGGAACTGCTACATGTGCGGGCATAAGGGCACCTNT  1980
           ||| ||||||||||||||||||||||||||||||||||||||||||||--------------T
Db  85861  CAAGCAGCCATTAAGGAAGACCCCTGGAACTGCTACATGTG--------------T  85902

Qy   1981  GGGCTGCCTGCGAAGACACGGGAAGACTGGGCCTTCTCGACTCCAGATGTTCTTTGCCAATAAC  2040
           ||||||| |||||||||||||||||||| ||||||||||||||||||||||||||||
Db  85903  GAGCTGCTGCGGAAGACAGGGGACTGGGCCTTCTCGACTCCAGAATGTTCTTTGCCAATAAC  85962

Qy   2041  CATGACCAGGAAATTGACCCCCAAAAGGTTTACCACACCTGCTGCACGTGCACGTGAGAGAGGAAG  2100
           || |||||||||||||||||||||||||| ||||||||||| ||||||||||||||||||||||
Db  85963  CACGACCAGGAAATTTGACCCCCGAAAGGTTTACCCACCTGCCCACGTGCACTGAGAGAGGAAG  86022

Qy   2101  CCCATCCGGCGTGCCTGTCTCTCTTTGATGGGGATTTGCTACAGGGGCTCCTGGGTGCTGAAGGAC  2160
           ||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  86023  CCCATCCAGGTGCCTATCCTCTATTTGATGGGAATTGCTACAGGGGCTCCTGGGTGCTGAAGGAC  86082

Qy   2161  CTGGGGCCATCCAAGTGGACCGGCCGCTACATTGCCTCCGAGGTGTCGTGAGGACTCCAATCACGGTG  2220
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  86083  CTGGGGCCATCCAAGTGGACCGGCCGCTACATTGCTTCCGAGGTGTCGATCACCCTG  86142

Qy   2221  GGCATGGTGCGCGCACCAGGGAAAGATCATCATGTACGTGGGGGACGTCCCGAGCGGTGCACACAG  2280
           |||| ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  86143  GGCATAGTGCGCGCCACCAGGGAAAGATCATGTTCGTCGGGGACGTCTCGCAGCATCACACAG  86202

Qy   2281  AAGCATATCCAGGAGTGTGGGGGGGGGCCCCATTCCGACCCTGGGTGATTGGAGGGCAGTGCTCCCTGCCAATGAC  2340
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  86203  AAGCATATCCAGGAGAGGGGGCCCCATACAATCTGGTGATTGGGGGTGCAGTGCTCCTTGCCAATGAC  86262

Qy   2341  CTCTCCATTGTCAAACCTGCCCCGCCCAAGGACACTTTATGAGGGGTGGTTACTGCCGGCTTCTTTT  2400
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  86263  CTCTCCATCATCAAACCCTGCCCCGGCAAAGGGACTTTACAAGGGCCACTGGCAGCCGCTCTTCTTT  86322

Qy   2401  GAGTTCTACCGGCCTCCTGCCATGATGCGCGCGGGCCCAAGGAGGGGGAGATGATCGCCCCCTTCTTC  2460
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  86323  GAGTTCTACCACCTCCTGCCATGATGCGCGGCCCAAGGAGGGGGAGGAGATCGCCCCCTTCTTC  86382

Qy   2461  TGGCTCTTTGACAATGTGGTGGCCCATGGGCGGTTAGTGACAAGAGGGGACATCTCGCGGATTT  2520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  86383  TGGCTCTTTGAGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  86442

Qy   2521  CTTGAGTCTAACCCCGTGATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGG  2575
           |           |
Db  86443  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG  86497
```

RESULT 9
AC112586/c

LOCUS   AC112586   176697 bp   DNA   linear   HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-28J22, *** SEQUENCING IN PROGRESS
            ***, 65 unordered pieces.
ACCESSION   AC112586
VERSION   AC112586.3 GI:21738467
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 176697)
  AUTHORS   Mozny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oraguzie,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,I.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 176697)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 176697)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 12, 2002 this sequence version replaced gi:20303171.
COMMENT   -------------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
          Center project name: GRAT
          Center clone name: CH230-28122
          -------------- Summary Statistics
          Sequencing vector: Plasmid;
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 121959 bases at least Q40
          Consensus quality: 126371 bases at least Q30
          Consensus quality: 130428 bases at least Q20
          ----
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 65 contigs. The true order of the pieces
          *   is not known and their order in this sequence record is
          *   arbitrary. Gaps between the contigs are represented as
          *   runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          *   as soon as it is available and the accession number will
          *   be preserved.
          *   1   1415: contig of 1415 bp in length
          *   1416   1515: gap of unknown length
          *   1516   2965: contig of 1450 bp in length
          *   2966   3065: gap of unknown length

*	3066	4228: contig of 1163 bp in length
*	4229	4328: gap of unknown length
*	4329	5385: contig of 1057 bp in length
*	5386	5485: gap of unknown length
*	5486	6670: contig of 1185 bp in length
*	6671	6770: gap of unknown length
*	6771	7813: contig of 1043 bp in length
*	7814	7913: gap of unknown length
*	7914	9673: contig of 1760 bp in length
*	9674	9773: gap of unknown length
*	9774	11470: contig of 1697 bp in length
*	11471	11570: gap of unknown length
*	11571	13162: contig of 1592 bp in length
*	13163	13262: gap of unknown length
*	13263	14504: contig of 1242 bp in length
*	14505	14604: gap of unknown length
*	14605	15974: contig of 1370 bp in length
*	15975	16074: gap of unknown length
*	16075	17396: contig of 1322 bp in length
*	17397	17496: gap of unknown length
*	17497	19323: contig of 1827 bp in length
*	19324	19423: gap of unknown length
*	19424	21040: contig of 1618 bp in length
*	21041	21140: gap of unknown length
*	21141	22418: contig of 1278 bp in length
*	22419	22518: gap of unknown length
*	22519	23941: contig of 1423 bp in length
*	23942	24041: gap of unknown length
*	24042	25332: contig of 1291 bp in length
*	25333	25433: gap of unknown length
*	25434	26530: contig of 1097 bp in length
*	26531	26630: gap of unknown length
*	26631	28270: contig of 1640 bp in length
*	28271	28370: gap of unknown length
*	28371	29970: contig of 1600 bp in length
*	29971	30070: gap of unknown length
*	30071	32265: contig of 2195 bp in length
*	32266	32365: gap of unknown length
*	32366	34179: contig of 1814 bp in length
*	34180	34279: gap of unknown length
*	34280	36392: contig of 2113 bp in length
*	36393	36492: gap of unknown length
*	36493	38552: contig of 2060 bp in length
*	38553	38652: gap of unknown length
*	38653	40932: contig of 2280 bp in length
*	40933	41032: gap of unknown length
*	41033	43021: contig of 1989 bp in length
*	43022	43121: gap of unknown length
*	43122	44977: contig of 1856 bp in length
*	44978	45077: gap of unknown length
*	45078	47358: contig of 2281 bp in length
*	47359	47458: gap of unknown length
*	47459	49554: contig of 2096 bp in length
*	49555	49654: gap of unknown length
*	49655	51876: contig of 2222 bp in length
*	51877	51976: gap of unknown length
*	51977	53794: contig of 1818 bp in length
*	53795	53894: gap of unknown length
*	53895	55634: contig of 1740 bp in length
*	55635	55734: gap of unknown length
*	55735	57445: contig of 1711 bp in length
*	57446	57545: gap of unknown length
*	57546	60210: contig of 2665 bp in length
*	60211	60310: gap of unknown length
*	60311	62174: contig of 1864 bp in length
*	62175	62274: gap of unknown length
*	62275	64285: contig of 2011 bp in length
*	64286	64385: gap of unknown length
*	64386	66764: contig of 2379 bp in length
*	66765	66864: gap of unknown length
*	66865	69398: contig of 2534 bp in length
*	69399	69498: gap of unknown length
*	69499	71451: contig of 1953 bp in length

RESULT 10
BC018214

Query Match   27.9%; Score 1171.2; DB 2; Length 176697;
Best Local Similarity 82.7%; Pred. No. 2.3e-268;
Matches 1317; Conservative 0; Mismatches 248; Indels 27; Gaps 3;

```
Db   23553  AATACGATGACGATGTGTACCAGTCCTTCTTGTGTACCACATCTGCTGTTGGGGGCGCGGAAGTGC 23494
Qy    1844  TCAATGTGTCGGGAACAACAACTGCTGCAGGTGCGTTTTGTGTCGTGCGATTCTCTTGG 1903
Db   23493  TCAATGTGTGGGGAACAACAACCGCTGCAGGTGCGTTTTGTGTCGGATTCTCTTGG 23434
Qy    1904  TGGGGCCAGGAGGCTGCTCAGGCAGCCATTAAGGAAGACCCCTGGAACTGCTACATGTGCG 1963
Db   23433  TGGGGGCCAGGGGGCTGCCCAAGCAGCCATTAAGGAAGACCCCTGAAACTGCTACATGTG-- 23376
Qy    1964  GGCATAAGGGCACCTATGGGGTGCTGCCGGAAGACGGGAAGACTGGCCTTCTCGACTTCCAGA 2023
Db   23375  ------------TGAGCTGCTGGGGACGGGAAGGAGGACTGGCCTTCTCGACTTCCACA 23332
Qy    2024  TGTTCTTTGCCAATAACCATCAGGAAGAATTGACCCCCCAAAGGTTTTACCACTGTGTGC 2083
Db   23331  TGTTTCTTTGCCAATAACCACCAGAAATTTGACCCCCCCAAAGGTTTACCACTGTGTGC 23272
Qy    2084  CAGCTGAGAAGAGGAGGAAGCCCATCCGGCGTGTCGTCCTCCTTGATGGGAAGACCTACAGGGC 2143
Db   23271  CAGCTGAGAAGAGGAGGAAACCCATCCACGTGCTGCATCTCTATTTGATGGGAATTGCTACAGGGC 23212
Qy    2204  AGGACTCCATCACACCCGTGGGGCATAGTAGTGCGGCACCAGGAAAGATCATGTTCGGTCGGGACG 2263
Db   23211  TCCTGGGTGCTGAAGGGAGCCTGGGCATCCAAGTGGACCCGTACATTGCCTTCCGAGGTGTGCG 23152
Qy    2204  AGGACTCCATCACACCCGTGGGGCATAGTGCGGCACCAGGAAAGATCATGTTCGGTCGGGACG 2263
Db   23151  AGGACTCCATCACCCGTGGGCATAGTGCGGCACCAGGAAAGATCATGTTCGGTCGGGACG 23092
Qy    2264  TCCGCGAGCGTCACACAGAAGCATATCCAGGAGTGTGGGGCCCCATTCGACCTGGTGATTGGAG 2323
Db   23091  TCTCGAGCATCACACAGAAGCATATCCAGGAGGGGGGCCCCATTACAATCTGGTGATTGGGG 23032
Qy    2324  GCAGTCCCTGCAAATGAACCTCTCCATTGTCAACCCTGCCCGGCCCGCCGGGACTTTATGAGGGTA 2383
Db   23031  GCAGTCCTTGCAAATGAACCTCTCCATCAACACCTGCCCGGCCCGCCAGGGGACTTTACAAGGGCA 22972
Qy    2384  CTGGCGCGGCCCTCTTCTTTGAGTTGCTCTACCGCCTCCCTGCATGATGCCGGGACCCCAGAGGAG 2443
Db   22971  CTGGCAGCCCGCCCTCTTCTTTGAGTTCTACCACCCTCCTGCATGATGCCGGGCCCCAGAGGAG 22912
Qy    2504  GGGACAGTCTGCGCGCGATTTCTTGACTCTAAACCCCGTGATCAGTTGACGGGCGCACCAGAAGTGTCTG 2563
Db   22911  AGGAATCGCGCCCTTCTTCTTGCTCTTTGAGGAAGGTGGTGGCCCATGGCCGTTAGTGACAAGA 22852
Qy    2564  CTGGCACACAGGGCCGGTTACTTCTGGGGTAACCTTCCTGGCATGAACAGGCCCTTTGGCAT 2623
Db   22791  CTCGCACACAGGGTCCGTTACTTCTGGGTAACCTTCCTGGCATGAACAGGCCCATTGGCAT 22732
Qy    2624  CCACTGTGAATGGATAAGCTGGAGCTGCAAGGTGCTCTGGAGCACGGCAGAATAGCCAAGT 2683
Db   22731  CCACTATGAATGGATAAGCTGGAGCTGCAAGGAGTGTCTGGAGCACAGCAGCCAGGT 22672
Qy    2684  TCAAGGCAAAGTGAGGAGGACCATTACCACCAGGTCAAACTCTATAAAGGCAAAGGGAAGGACCAGC 2743
Db   22671  TCAGCAAAGTGAGGAGGACCATTACCACCAGGTCAAACTCCGTGAAATGGGAAAGG 22612
Qy    2744  ATTTCCCCCGTGTCTTCATGAACCAGGAAGAGGAGACATCGTGGTGGTCACTGAAATGGACACC 2803
Db   22611  ATTTCCCTGTCTTCATGAATGAGAAGGAAGACATCCTCTGGTGCACTGAAATGGGAAACAG 22552
Qy    2804  TGTTTGGCTTCCCCGTCCCACTACACAGACGTC 2835
Db   22551  TGTTTGGCTTCCCCGTCCCACTACACAGANAGCC 22520
Qy    1784  AGTATGACGACGATGGGGTACCAGTCCTATTTGCACCATCTGCTGTGGGGGCGCGGAAGTGC 1843
Db    1724  CACTCTTCATTGGGTGGAAATGTGCCCACAACTGTAAGAAACTGCTTCCTTGGAGTGTGCTTACC 1783
```

```
LOCUS       BC018214              1758 bp    mRNA    linear   PRI 06-DEC-2001
DEFINITION  Homo sapiens, clone IMAGE:3862699, mRNA, partial cds.
ACCESSION   BC018214
VERSION     BC018214.1  GI:17390481
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1758)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 21 Row: h Column: 23.
FEATURES             Location/Qualifiers
     source          1..1758
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3862699"
                     /clone_lib="NIH_MGC_66"
                     /tissue_type="Ovary, adenocarcinoma"
                     /lab_host="DH10B"
                     /note="vector: pCMV-SPORT6"
     CDS             <1..859
                     /codon_start=2
                     /product="Unknown (protein for IMAGE:3862699)"
                     /protein_id="AAH18214.1"
                     /db_xref="GI:17390482"
                     /translation="AEKRKPIRVLSLFPGIATGLLVLKDLGIQVDRVIASEVCEDSIT
                     VGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGR
                     LFFEYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSA
                     AHRARYFWGNLPGHNRPLASTVNDKLELQEELEHGRIAKFSKVRTITTRSNSIKQGKD
                     QHFPVFMNKEDILWCTEMERVPGFPVHTDVSNMSRLARQRLLGRSWSVPVIRHLFA
                     PLKKYFACV"
BASE COUNT      513 a    376 c    428 g    441 t
ORIGIN

Query Match      27.8%;  Score 1164.8;  DB 9;  Length 1758;
Best Local Similarity   86.2%;  Pred. No. 4.9e-267;
Matches 1483;  Conservative   0;  Mismatches 172;  Indels  65;  Gaps  15;
```

```
Qy 2085 AGCTGAGAAGAGGAAGAGCCCATCCGGCGGTCGTCTCTCTTTGATGGGATTGCTACAGGGCT 2144
        ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    1 AGCTGAGAAGAGGAAGCCCATCCGGCGGTCGTCTCTCTTTGATGGGATTGCTACAGGGCT 60

Qy 2145 CCTGGGTGCTGAAGGAACCTGGGCATCCAAGTGGACCCGCCGCTACAGGTGGTGTTGTGA 2204
        ||||||||||||||||||||||||||||||||| | |||||||||||||||||||||||
Db   61 CCTGGGTGCTGAAGGACTTGGGCATTCAGGTGGACCCGCTACAGGTGGTGTTGTGA 120

Qy 2205 GGACTCCATCACGGTGGGCATGGTGCGGCACCAGGGAAAAGATCATTGTACGTCGGGACGT 2264
        |||||||||||||||||||||||||||||||||| ||||| |||||||||||||||||||
Db  121 GGACTCCATCACGGTGGGCATGGTGCGGCACCAGGGAAGATCATTGTACGTCGGGACGT 180

Qy 2265 CCGCAGCGCCTCACACAGAAGCAATATCCAGGAGTGGGGGCCCATTCGACCTGGTGATTGGAGG 2324
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CCGCAGCGCCTCACACAGAAGCAATATCCAGGAGTGGGGGCCCATTCGACCTGGTGATTGGGGG 240

Qy 2325 CAGTCCCTGCAATGACCTCTCCATTGTCAACCCTGCCCGCAAGGGACTTTATGAGGGTAC 2384
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  241 CAGTCCCTGCAATGACCTCTCCATCGTCAACCCTGCCCGCAAGGGACTTTATGAGGGCAC 300

Qy 2385 TGGCCGGCGCCTCTTCTTCTTTGAGTTCTACCGCCTCCTGCACGATGCGCGGCCCAAGGA 2444
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TGGCCGGCGTCTTCTTCTTTGAGTTCTACCGCCTCCTGCACGATGCGCGGCCCAAGGA 360

Qy 2445 TGATCGGCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAG 2504
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TGATCGGCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGAG 420

Qy 2505 GGACATCTCGCGATTTCTTGAGTCTAACCCCGTGATGATTGACGCCAAAGAAGTGTCTGC 2564
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GGACATCTCGCGATTTCTGGAGTCCAACCCCGTGATGATTGATGCCAAAGAAGTGTCAGC 480

Qy 2565 TGCACAGAGGGCCCGTTACTTCTGGGGGTAACCTTCCTGGCATGAACAGGGCCTTGGCATC 2624
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TGCACACAGGGCCCGCTACTTCTGGGGGTAACCTTCCCGGTATGAACAGGGCCGTTGGCATC 540

Qy 2625 CACTGTGAATGATAAAGCTGCAAGAGTGTCTGGAGCACGGCCAGAATAGCCAAAGTT 2684
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CACTGTGAATGATAAAGCTGCAAGAGTGTCTGGAGCCAGAATAGCCAAAGTT 600

Qy 2685 CAGCAAAGTGAGGACCATTACCACCAGGTCAAACTCTATAAAGCAGGGCCAAAGACCAGCA 2744
        |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  601 CAGCAAAGTGAGGACCATTACTACCAGGTCAAACTCCATAAAAGCAGGGCAAAGACCAGCA 660

Qy 2745 TTTCCCCGTCTTCATGAACGAGGAGGAGGACATCCTGTGGTGCACTGAAATGGAAAGGGT 2804
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TTTTCCTGTCTTCATGAATGAGGAGGACATTCTTTGGTGCACTGAAATGGAAAGGGT 720

Qy 2805 GTTTGGCTTCCCCGGTCCCACTACACAGACGTCTCCAACATGAGCCCGCTTGGCGAGGCAGAG 2864
        ||||| ||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  721 ATTTTGGTTTCCCAGTCCCACTATACTGACGTTCCAACATGAGCCCGCTTGGCGAGGCAGAG 780

Qy 2865 ACTGCTCGGGCGCCGATCGTGGGACCGGTGCCGGTCATCCGCCACCTCTTCGCTCCGGCTGAAGGA 2924
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 ACTGCTCGGGCCGGTCATCGAGGCGGTGCCCAGTCATCGCCGCCACCTCTTCGCTCCGGCTGAAGGA 840

Qy 2925 ATATTTTTGCCTTGTGTGTGTAAAGGGCACATGGGGGCAAACTGAAGTAGTGATGATAAAAAAGTTT 2984
        |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  841 GTAATTTGCGTTGTGTGTGTAAGGGACCATGGGGGCCAAAACTGAGGTAGGGA----CACAAAGTT 896

Qy 2985 AAACAAACAAAACAAAACAAAACAAAATAAAACAACAAGAGAGGAGGACACG 3044
        ||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  897 AAACAAACAAAACAAAACAAAAACACAAAAACAT-----AATAAAAACACCAAGAACATGAGGATGG 951

Qy 3045 AGAAAAGT-TCAGCACCCAGAGAGAGAGAGAAAAAGGAAATTTAAAGC--AAACCACAGAGGAGGA 3101
        || ||||| |||||||||||||||||||||| |||||||||||||||| ||||||||||||||||
Db  952 AGAGAAGTATCAGCACCCAGAGAGAGAGAAAAAGGAATTTAAAACAAAACCACACAGAGGCGGA 1011

Qy 3102 AAACGCCGGAGGGCTTGGCCTTGCAAAAGGGGTTGGACATCATCTCTGAGTTTTCAATGT 3161
        || ||||||| ||| |||||||||||||||||||||||||||||||||||||||||||||
Db 1012 AATACCGGAGGGCTTTGCCTCGGCAAAAGGGGTTGGACAATCATCTCCTGATTTTTCAATGT 1071

Qy 3162 TAACCTTCAGTCCTATCTATAAAAAGCAAAAATAGGC--CCCTCCCCTTCTTCCCC---TCCG 3217
        |||||||| |||||||||||| ||||||||||||||| |||||||||||||||||   ||||
Db 1072 TATTCTTCAGTCCTATTTATTAAAAACAAAATCCAAGCTCCCTCCCCTTCCTCCCCCTTCCCTT 1131

Qy 3218 TCCTAGGAGGCAGGCGAAACTTTTTTGTTTTTTTCTTTTCAGAGGGGGTTTTGTGTTTTGG 3277
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1132 TTTTTTCGGTCAGACCTTTTATTTTTTTCTTTTCAGAGGGGTTTTCTGTTTTGTTTTGG 1190

Qy 3278 GTTTTGTTTCTTGCTGTGACTGAGACAAGAAAGAGGAGTTATTGCAGC--AAAATCAGTAACAAC 3336
        | ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db 1191 G-TTTTGTTTTCTTGCTGTGACTGAAAACAAGAAAGAGGTTATTGCAGCAAAAATCAGTAACAAA 1249

Qy 3337 AAAAAGTAGAAAATGCCTTGGAGGGAAAGG--------GAGAGAGGGAAAAATTCTATAA 3387
        ||||||||||||||||||||||||||||||        ||||||||||||||||||||||
Db 1250 AAAAAGTAGAAAATGCCTTGGAGGGAAAGGGAGAGAGGGAAAAATTCTATAA 1387
```

RESULT 11
AC103454
AC103454/c

LOCUS   AC103454   117264 bp   DNA   linear   HTG   12-JUL-2002
DEFINITION   Rattus norvegicus clone CH230-130J17, *** SEQUENCING IN PROGRESS
             ***, 60 unordered pieces.
ACCESSION   AC103454
VERSION   AC103454.3 GI:21728739
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 117264)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
             Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
             Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Rolle,S., Saverly,G.,
             Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
             Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
             Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
             Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
             Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,P., Wang,Q.,
             Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
             Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
             Wu,C., Wu,Y., Wu,F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 117264)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 117264)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jul 11, 2002 this sequence version replaced gi:1973284.
             --------------------- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             --------------------- Project Information
             Center project name: GHPN
             Center clone name: CH230-130J17
             --------------------- Summary Statistics
             Sequencing vector: Plasmid;
             Chemistry: Dye-terminator Big Dye: 100% of reads
             Assembly program: Phrap; version 0.990329
             Consensus quality: 54254 bases at least Q40
             Consensus quality: 56889 bases at least Q30
             Consensus quality: 59339 bases at least Q20
             * NOTE: Estimated insert size may differ from sequence length
             * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 60 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             * ---------------------
             *   1: contig of 1343 bp in length
             *   1343: gap of unknown length
             *   1344: contig of 1139 bp in length
             *   2582: gap of unknown length
             *   2583: contig of 1035 bp in length
             *   2683: gap of unknown length
             *   2683: contig of 1095 bp in length
             *   3717: gap of unknown length
             *   3718: contig of 1337 bp in length
             *   4913: gap of unknown length
             *   4912: contig of 1337 bp in length
             *   5012: gap of unknown length
             *   5013: contig of 1425 bp in length
             *   6350: gap of unknown length
             *   6449: contig of 1076 bp in length
             *   6450: gap of unknown length
             *   7874: contig of 1076 bp in length
             *   7875: gap of unknown length
             *   7975: contig of 1076 bp in length
             *   9050: gap of unknown length
             *   9051: contig of 1465 bp in length
             *   9151: gap of unknown length
             *   10615: contig of 1543 bp in length
             *   10616: gap of unknown length
             *   10716: contig of 1543 bp in length
             *   12258: gap of unknown length
             *   12259: contig of 1317 bp in length
             *   12359: gap of unknown length
             *   13675: contig of 1317 bp in length

Db   1250 AAATAGTAACAATACCTTGCAGAGAGAAAGGTGGGAGAGGAAAAAAGGAAAATTCTATAG 1309
          ||| |||| |||| || |||||||||||||||||||||||||||||||||||||||||
Qy   3388 AAACTTAAAAATATTGGGTTTTTTTTTTTTTTCCTTT------------------ 3422

Db   1310 AAATCTATATATTGGGTTGTTTTTTTTTTTTTGTTTTTTTTTTTTTTTGGGTTTTTT 1369
          |||| | || |||||||| || |||||||||||||||||||||||||||||||||
Qy   3423 ------TCTATATATCTCTTTGGTTGTGTCTCTAGCCTGTTCGATCAGATAGGAGCACAAACA--G 3474

Db   1370 TTTTACTATATATATCTTTTTTGTTGTCTCTAGCCTGATCAGATCAGATAGGAGCACAAGCAGG 1429
          | ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1430 GACGGAAAGAGAGAGACACTCAGGCGGCGAGCAATTCCCTCCAGCCACTGAGCTGTCGTGC 1489
          | | || ||| | |||| || |||||||||||||||||||||||||||||||||||
Qy   3475 GAAGAGAATAGAGAGACCTCGGAGGCAG-AGTCTCCTCCACCCCCGAGCCAGTCTCAA 3533

Db   1490 CAGCACCCATTCCTGGTCATGCATAAAAACAGAACCCAACTAGTGCAGGGCGGCTCAGAGAACA 1549
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3534 CAGCACCATTCCTGGTCACGCAAAAACAAGCAGGCGTG||||CAGGGCGGCTCAGAGAACA 3593

Db   1550 ACACAAGACA-TTTTCTTCACAGTATTTCAGGTGCCTACCACACAGGAAAAACCTTGAAGAAA 1608
          ||| | ||| | |||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3594 CCACACCAGACACTTTCTACACAGTATTTCAGGTGCCTACCACACAGGAAAACCTTGAAGAAA 3653

Db   1609 ATCAGTTTCTAGAAGCCGCTGTTPACCTCTTGTTTPACCAGTTTATATATATATGATAGATAT 1668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3654 ACCAGTTTCTAGAAGCCGCGCTGTTTTPACCTCTTTPACAGTTTATATATATATATGATAGATAT 3713

Db   1669 GAG--ATATATATAAAAAGGTAGTACTGTTAACTACTACTGTACA 1706
          ||| |||||||||||||||||||||||||||||||||||||||
Qy   3714 GAGATATATATATAAAAAGGTACTGTTAACTACTACTGTACA 3753

REFERENCE
  AUTHORS   Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
             Barbaria,J., Bimage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Deaerich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
             Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
             Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
             Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
             Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
             Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
             Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
             Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
             Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
             Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

```
* 13676    13775: gap of unknown length
* 15423    15523: contig of 1648 bp in length
* 15424    15523: gap of unknown length
* 15524    16891: contig of 1368 bp in length
* 16892    16991: gap of unknown length
* 16992    18433: contig of 1442 bp in length
* 18434    18533: gap of unknown length
* 18534    19859: contig of 1326 bp in length
* 19860    19959: gap of unknown length
* 19960    21010: contig of 1051 bp in length
* 21011    21110: gap of unknown length
* 21111    22710: contig of 1600 bp in length
* 22711    22810: gap of unknown length
* 22811    24047: contig of 1237 bp in length
* 24048    24147: gap of unknown length
* 24148    25330: contig of 1183 bp in length
* 25331    25430: gap of unknown length
* 25431    26558: contig of 1128 bp in length
* 26559    26658: gap of unknown length
* 26659    28182: contig of 1524 bp in length
* 28183    28282: gap of unknown length
* 28283    29368: contig of 1086 bp in length
* 29369    29468: gap of unknown length
* 29469    30746: contig of 1278 bp in length
* 30747    30846: gap of unknown length
* 30847    32194: contig of 1348 bp in length
* 32195    32294: gap of unknown length
* 32295    33410: contig of 1116 bp in length
* 33411    33510: gap of unknown length
* 33511    35012: contig of 1502 bp in length
* 35013    35112: gap of unknown length
* 35113    36244: contig of 1132 bp in length
* 36245    36344: gap of unknown length
* 36345    37377: contig of 1033 bp in length
* 37378    37477: gap of unknown length
* 37478    39062: contig of 1585 bp in length
* 39063    39162: gap of unknown length
* 39163    40659: contig of 1497 bp in length
* 40660    40759: gap of unknown length
* 40760    41926: contig of 1167 bp in length
* 41927    42026: gap of unknown length
* 42027    43282: contig of 1256 bp in length
* 43283    43382: gap of unknown length
* 43383    44987: contig of 1605 bp in length
* 44988    45087: gap of unknown length
* 45088    46483: contig of 1396 bp in length
* 46484    46583: gap of unknown length
* 46584    47680: contig of 1097 bp in length
* 47681    47780: gap of unknown length
* 47781    49170: contig of 1390 bp in length
* 49171    49270: gap of unknown length
* 49271    51444: contig of 2174 bp in length
* 51445    51544: gap of unknown length
* 51545    52836: contig of 1292 bp in length
* 52837    52936: gap of unknown length
* 52937    55045: contig of 2109 bp in length
* 55046    55145: gap of unknown length
* 55146    56691: contig of 1546 bp in length
* 56692    56791: gap of unknown length
* 56792    58443: contig of 1652 bp in length
* 58444    58543: gap of unknown length
* 58544    60538: contig of 1995 bp in length
* 60539    60638: gap of unknown length
* 60639    63571: contig of 2933 bp in length
* 63572    63671: gap of unknown length
* 63672    65119: contig of 1448 bp in length
* 65120    65219: gap of unknown length
* 65220    66996: contig of 1777 bp in length
* 66997    67096: gap of unknown length
* 67097    69660: contig of 2564 bp in length
* 69661    69760: gap of unknown length
* 69761    72469: contig of 2709 bp in length
* 72470    72569: gap of unknown length
```

```
* 72570    75117: contig of 2548 bp in length
* 75118    75217: gap of unknown length
* 75218    76930: contig of 1713 bp in length
* 76931    77030: gap of unknown length
* 77031    78952: contig of 1922 bp in length
* 78953    79052: gap of unknown length
* 79053    82025: contig of 2973 bp in length
* 82026    82125: gap of unknown length
* 82126    83473: contig of 1348 bp in length
* 83474    83573: gap of unknown length
* 83574    86343: contig of 2770 bp in length
* 86344    86443: gap of unknown length
* 86444    90103: contig of 3660 bp in length
* 90104    90203: gap of unknown length
* 90204    95081: contig of 4878 bp in length
```

```
Query Match          25.2%;     Score 1055.2;   Pred. No. 1.1e-240;   DB 2;   Length 117264;
Best Local Similarity 90.4%;
Matches 1304;   Conservative 0;   Mismatches 70;   Indels 68;   Gaps 14;
```

```
Qy    2798 AAAGGGTGTTTGGGCTTCCCCGGTCCCACTCACACAGACGTCTCCAACATGAGCCGCGCTTGGCGA 2857
           +|||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  105256 ACAGGGTGTTTGGGCTTCCCTGTCCCACTCACACAGACGTCTCCAACATGAGCCGCGCTTGGCGA 105207

Qy    2858 GGCAGAGAGACTGCTGTGGGCCGATCGTGGAGCGGTGCCCGGTCATCCGGCCACCTCTTCGCTCCGC 2917
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105206 GGCAGAGAGACTGCTGTGGGCCGATCGTGGAGCGGTGCCAGTCATCCGGCCACCTCTTCGCTCCGC 105147

Qy    2918 TGAAGGAAATATTTGCTTGTCGTGTGAAGGACATGGGGCGCAAACTGAAGTAGTGAATGATGAATAA 2977
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105146 TGAAGGAAATATTTTGCTTGTCGTGTGTAAGGGACACATGGGGGCAAAACTGAAGTAGTGAATGATAA 105087

Qy    2978 AAAAGTTAAACAAACAAACAATAAAAACAAAAACAATAAAAACACCAAGAAACGAG 3037
           ||  ||||||||||||||||||||||||  |||||||||||||||||||||||||
Db  105086 CCAAGTTTAAACAAACAAAAAAAACAA----------AACACAATAAAACACCAAGAAACCAG 105037

Qy    3038 AGGACGGAGAAAAGTTCAGCACCCAGAGAGAAAAAGGAATTTAAAGCAAAACCACAGAGG 3097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105036 AGGACGGAGAAAAGTTCAGCACCCAGAGAGAAAAAGGAATTTAAAGCAAAACCACAGAGG 104977

Qy    3098 AGGAAAAACGCCGGAGGGGCTTGGCCTTGCAAAAAGGGGTTGGACACATCATCTCCTGAGTTTTCA 3157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104976 AGG-AAAATGCCCGGAGGGGCTTGGCCTTGCAGAAGGGGTTGGACATCATCTCTTGAGTTTTCA 104918

Qy    3158 ATGTTAACCTTCAGTCCTATCTAAAAAGCAAAAATAGGCCCCTCCCCTTCTTTCCCCTCCGG 3217
           | |||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  104917 AIGTTCATCTTCAGTCCTACTAAACGAAGGATAGGCCCCTCCCCTTCTTTCCCCTCCGG 104858

Qy    3218 TCCTAGGAGGGGGAACTTTTTTGTTTTCTACTC----TTTTTCAGAGAGGGGGTTTTTCTGTTTGTTT 3275
           ||||||||||||||||||||||||||||||||||    ||  |||||||||||||||||||||||||||
Db  104797 GGGTTTTTGTTTCTTGCCTGTGACTCAAAACAGAGAGAGTTAATTGCAGCAAAAATCAGTAACA 104738

Qy    3276 GGGTTTTTTGTTTCTTGCTGTGACTCGAAACAACAAGAGAGTT-ATTGCAGCAAAATCAGTAACA 3334
           ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  104857 TCCTAGGAGGCGCAAACCTTTTTGTTTTCTACTCTTTTTTCAGAGGGGGTTTTTCTGTTTGTTT 104798

Qy    3335 ACAAAAAGTAGAAAATGCCCTTGGAGAGGAAAGGGGAGGAGGAAAAATTCTATAAAAACTTA 3394
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104737 ACAAAAAGT---AATGCCCTTGGAGAGGAAAGGGGAGGAGGAAAAATTCTATAAAAAACTTA 104681

Qy    3395 AAATATTGGTTTTTTTTTTTTTTTTTTTTCCTTTTTCTATATATCTCTTTGGTTGTGTCTCTAGCCTGA 3454
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104680 AAATATTG-------TTTTTTTTTTTTTTTCCTTTTTCTATATATCTCTTTGGTTGTGTCTCTAGCCTGA 104624

Qy    3455 TCAGATAGGAGGCACAAAACAGGAAAGAAATAGAGACCCTCGGGAGGCAGCAGAGTCTCCTCTCCC 3514
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104623 TCAGATAGGAGGCACAAAACAGGAAACAGAAATAGAGACCCTCGGGAGGCAGCAGAGTCTCCTCTCGC 104564

Qy    3515 ACCCCCGAGCGAGTCTCAACAGCACCATTCCTGGTCATGCAAAAACAGAAACCCAAACTAGCA 3574
           | ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3515 ACCCCCGAGCGAGTCTCAACAGCACCATTCCTGGTCATGCAAAAACAGAAACCCAAACTAGCA 3574

Qy    3575 GCAGGGCGCTGAGAGAACACCACCACGACACTTTCTACAGTATTTTCAGGTGCGCTGCCTACCAC 3634
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104563 AGCCCCCGAGCGGGTCTCAACAGCACCATCCCGTCATGCAGAACAGAAACCCAAACTAGCA 104504
```

```
Db  104503  GCAGGGCGCTGAGAATACCACCACCAGACAC-TTTTCTACAGTATTTCAGGTGCCTACCAC  104445
            ||||||||||||| |||   |||                 |||||||||||||||||||||
Qy    3635  ACAGGAAACCTTGAAGAAAACCAGTTTCTAGAAGCCGCTGTTACCCTCTTGTTTTACAGTTT  3694
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104444  ACAGGAAACCTTGAAGAAAACCAGTTTCTAGAAGCCGCTGTTACCCTCTTGTTTTACAGTTT  104385

Qy    3695  ATATATATATGATAGATATGAGTATATATATATATATATATAAAGGTACTCTTAACTACTGTACAT  3754
            ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104384  ATATATATACGATAGATATGAG--ATATATATATATAAAAAGGTTACTTAACTACTGTACAT  104327

Qy    3755  CCCGGACTTCATAAATGGTGGCTTTCAAAAACACAGCGAGATGAGCAAAAGACATCAGGTTCCGCCT  3814
            |||   ||||||||||||| |||||||||||||||||||||||||||  ||| |||| |||
Db  104326  CCCAACTTCATAAATGGTGCTTTCAAAAACAGCGAGATGAGCAAAAACATCAAACCTCCACCT  104267

Qy    3815  GGCCCTCTGTCAAAACGGGTTTCAGCCCACAGGATGGGGAGAGGGGAGGGCAGCTGGAGG-----  3869
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104266  GGCCCTCTGTCAAAAGGGGTTTCAGCCCAGGATGGGG-GAGGAAGCAGCTAGAGGGGTGTC  104208

Qy    3870  -----------------------------GGGTTTAACAAAACTGAAGGATGACCCATAT-  3900
                                         || ||||||| || | || ||||
Db  104207  GTTTTTTTTTTTTTTTTTTTTTTTTTTGTCCCTTCTTGGTTTTTAACAAACCAAAGGAGGANCCATATT  104148

Qy    3901  ---CACCCCCACCCCCTGCCCCTCTAGCTTCACCTGCCAAA-AGGGGCTCAGCTGA  3956
               ||||  ||||||||||| |||||||||||||||||||||   ||||||||||||
Db  104147  GCCCCCACCCCCACCCATGCCCTAGCTTCTCCTGCCAAAAGTGGGGCTCAGCTGA  104088

Qy    3957  GGTGGGTCGGACCCTGGGGGAAGCTGAGTGTGGAATTTATCCAGACTCGGCGTGCAATAACCT  4016
            |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  104087  GGTGGGTCGGAACCTGGGGGAAGCGGAGTGTGGAATTTATCCAGACTCGGCGTGCAATAACCT  104028

Qy    4017  TAGAATATGAAATCTAAAAATGACTGCCTCAGAAAAAATGGCTTGAGAAAACA-----TTCTC  4071
            |||||||||||| |||||||||||||||||||||||||||| |||||||||||     |||||
Db  104027  TTGAATATGAAATCTAAAAACGACTGCCTCAGAAAAAATGGCTTAAAAAAAAAAGTTCTC  103968

Qy    4072  CCTGATTTTGAAATTCGTCAGCCACGTTGAAGG-CCCCTTGTGTGGGATCAGAAATATTCCAG  4130
            |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db  104027  CCTGATTTTGAAATCTGTCAGCCACATTGAAGGCCCCCTTGTGTGGGATAAGAAATAATCCAG  103908

Qy    4131  AGTGGAGGGAAAGTGACCCGCCCATTAACCCCNCTGGAGCAAATAAAAAAAACATACAAAAT  4190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103907  AGTGAGGGAAAGTGATCCGCCATTAACCCCACCTGGAGCAAATAAAAAAACATACAAAAC  103848

Qy    4191  GT  4192
            ||
Db  103847  GT  103846
```

RESULT 12

LOCUS   AC009474   168651 bp   DNA   linear   PRI 08-NOV-2000
DEFINITION Homo sapiens BAC clone RP11-179G23 from 2, complete sequence.
ACCESSION   AC009474
VERSION   AC009474.4 GI:8748849
KEYWORDS   HTG.
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168651)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2 (bases 1 to 168651)
  AUTHORS   Mulvane.E., Stoneking,T., LaPlant,Y. and McDill,B.
  TITLE   The sequence of Homo sapiens BAC clone RP11-179G23
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 168651)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 168651)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 168651)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Jun 27, 2000 this sequence version replaced gi:7630880.
            -------------- Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
               Center project name: H_NH0179G23

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR: pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-547F18; the clone sequenced
            to the right is RP11-444B4, 200 base pair overlap. Actual end of
            this clone is at base position 18711 of RP11-444B4.
FEATURES   Location/Qualifiers
     source   1..168651
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-179G23"
                     /clone_lib="RPCI-11"
     repeat_region   13..497
                     /rpt_family="L2"
     repeat_region   616..683
                     /rpt_family="L2"
     repeat_region   1615..1815
                     /rpt_family="MIR"
     repeat_region   2577..2748
                     /rpt_family="MIR"
     repeat_region   2954..3144
                     /rpt_family="MIR"
```

```
repeat_region   3149..3465      /rpt_family="MER1_type"
repeat_region   3510..3567      /rpt_family="L1"
repeat_region   3569..3760      /rpt_family="MER1_type"
repeat_region   3827..4198      /rpt_family="MER1_type"
repeat_region   5656..6129      /rpt_family="MER1_type"
repeat_region   6205..6316      /rpt_family="MaLR"
repeat_region   6428..6686      /rpt_family="L2"
repeat_region   7247..7437      /rpt_family="L1"
repeat_region   7819..7868      /rpt_family="MER53"
repeat_region   9304..9469      /rpt_family="CR1"
repeat_region   10317..10439    /rpt_family="Alu"
repeat_region   12219..12455    /rpt_family="MIR"
repeat_region   12467..13537    /rpt_family="Alu"
repeat_region   13553..14453    /rpt_family="L1"
repeat_region   14853..15123    /rpt_family="L1"
misc_feature    16205..16345    /rpt_family="MER2_type"
misc_feature    16209..16345    /note="similar to EST AA338793 (NID:g1991103)"
misc_feature    16210..16345    /note="similar to EST AI123298 (NID:g3539064) ov48c08.x1"
misc_feature    16513..16676    /note="similar to EST AL037015 (NID:g5928008)"
repeat_region   16970..17198    /rpt_family="ERVL"
repeat_region   17199..17548    /rpt_family="L1"
repeat_region   17575..18261    /rpt_family="MER1_type"
repeat_region   19409..19531    /rpt_family="L1"
repeat_region   19540..19659    /rpt_family="L1"
repeat_region   25486..25822    /rpt_family="Alu"
misc_feature    26074..26366    /note="similar to EST AI928247 (NID:g5664211) wo95h11.x1"
repeat_region   26384..26542    /rpt_family="Alu"
repeat_region   26569..27058    /rpt_family="MIR"
repeat_region   27140..27647    /rpt_family="L2"
repeat_region   30824..31150    /rpt_family="L2"
repeat_region   32262..32525    /rpt_family="MaLR"
repeat_region   32568..32770    /rpt_family="CR1"
repeat_region   32775..33032    /rpt_family="MER1_type"
repeat_region   33172..33418    /rpt_family="CR1"
repeat_region   34930..35230    /rpt_family="MIR"
repeat_region   35302..35555    /rpt_family="Alu"
repeat_region   ...35555        /rpt_family="L2"

repeat_region   35975..36081    /rpt_family="L2"
repeat_region   36364..36515    /rpt_family="L2"
repeat_region   37266..37340    /rpt_family="Alu"
repeat_region   37400..37440    /rpt_family="L2"
repeat_region   37615..37715    /rpt_family="L2"
repeat_region   38006..38177    /rpt_family="MIR"
repeat_region   43171..43204    /rpt_family="MER1_type"
repeat_region   44777..44926    /rpt_family="U2"
repeat_region   45905..46110    /rpt_family="MIR"
repeat_region   48479..48638    /rpt_family="MIR"
repeat_region   48914..49033    /rpt_family="MER1_type"
repeat_region   49069..49337    /rpt_family="MIR"
repeat_region   49340..49521    /rpt_family="MER1_type"
repeat_region   50377..50498    /rpt_family="Alu"
repeat_region   56393..56545    /rpt_family="MIR"
repeat_region   58109..58420    /rpt_family="MIR"
```

```
Query Match     24.2%;  Score 1015.8;  DB 9;  Length 168651;
Best Local Similarity  68.7%;   Pred: No. 3.1e-231;
Matches 1881;  Conservative 0;  Mismatches 482;  Indels 376;  Gaps 20;

Qy    843 GGCTGAGAAGAAA-GCCAAAGGTAATTCGCAGTAAATGAAATGCTGTGAAGAGAAACAGGCCCT    901
          ||||||||||||| ||| |||| | |||||||||||||||||||||| ||| ||||| |||| ||
Db  74854 GGCTGAGAAGAAGAAAGGGCCGAGGTCACTCGCAGTAAATGAAATGCTATGAAAGAAAACTAGGGCT  74913

Qy    902 CTGGAGAGTCTCAGAAGGTGGAGGAGGGCCAGCCCTCCTCGTCGTGCAGCAGCGCCACCGGACC    961
          | | |||||| | ||||||||||||||||||||||||||||||||||||  |||| |||
Db  74914 CAGAGGAGTCTGCAAAGGTGGAGGAGGGCCAGCCCTCCTCGTCGTGCGCGCAGCCACC          74973

Qy    962 CTGCTTCTCCGACTGTGGCCCACCCACCCCTGAGCCCAGTAGGAGGGGGATGCTGGGGACAAGA   1021
          | || | ||| ||||| ||  |||| ||| |||||| ||   || |||   || ||| ||||| |
Db  74974 CCACATCCCCCACTGTGTCGTTACCACACCGTGAGCCTGTGGGGACTGAGGCTGGGAACAAGT    75033

Qy   1022 ATGCTACCAAAAGCAGCCCGACGATGAGCCTGAGTAATGAGGATGGCCCGGGGCTTGGCCATTG   1081
          |   ||| |||||||| | ||||||  ||  |||||  ||||| | ||||||| |||| ||| ||
Db  75034 A---CATCAAAGCAGGTGACGATGAACCAGAGTACAAGGACAGCCTGGGGCTTTTGGCAATG     75090

Qy   1082 GAGAGCTGGTCGTGGGGGAGAAACTTCGGGGGCTTCTCCTGGTGGCCCAGGCCCGAATTGTGTCTT  1141
          | |||||||| |||  || |||||| |  |||| |||||||||    |||| ||| ||||||||||
Db  75091 GGGAGCTGGTGTGTGAGGAAAAACTGCCAGGGCCTTCTCCTGGT----GCCCGCATTGTGTCTT   75145

Qy   1142 GGTGGATGAACAGGCCGGAGCCGGAGCAGCTGAAGGGCACTCGCTGGGGTCATGTGGTTCGGGAG  1201
          |||||||| ||||||| | |||  ||||||||||||||||  | ||| ||||||||  |  |||
Db  75146 GGTGGATGACACAGGCCAAAGCCCGAGCAGCTGAAGGGCACCCACTGTGTCATGTGGGCTTGGAG   75205

Qy   1202 ATGGCAAGTTCTCAGTGGTGGTGTGGAGAAGCTCATGCCCCTGAGCCTCCTTCTCCGAGTG      1261
          | || |||||||||||| ||  ||||  |||| |  | ||||||| |||||||  | |||||
Db  75206 A-GGTGAGTTCTCAGTGCTGCTGTGTAGAGAAGCCTGATACCACTGAGCCTCCTTTTGCAGTG    75264

Qy   1262 CATTCCACCAGGCCCACC-------TACAACAAGCAGCACCCCATGTACCCCGCCAAAGGCCATCTACGA  1316
          |||||||||| ||||||       ||| | |||||||| ||||||||||| || ||  ||||| | ||
Db  75265 TGTTCCACCAGACCCACCTTATAATACACAACAAGCAGCCCCATGTACCACAAAGCCATCACGA    75324

Qy   1317 AGTCCTCCAGGTGGCCCAGCCCGTGCCGGAAGCTGTTTCCAGCTTGCCCATGACAGTGA      1376
          |||||| | ||||  |||||| |||  | ||| ||| | |||||  |||| ||||||||
Db  75325 AGTCCTGCCGGTGCCCAGCCAG-TGTACAGGCGAGCTGTTCTCAGCGGAGCCAATGACAGTGA   75383

Qy   1377 TGAAAGTGACAGTGGCAAGGCTGTGGAAGTGCAGAACAAGCAGGAGATGAAATGGGCCCT     1436
```

```
RESULT 13
LOCUS       AC112040/c
ACCESSION   AC112040/c
DEFINITION  Rattus norvegicus clone CH230-61E10, *** SEQUENCING IN PROGRESS
            ***, 59 unordered pieces.
LOCUS       AC112040    119630 bp    DNA    linear    HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-61E10, *** SEQUENCING IN PROGRESS
            ***, 59 unordered pieces.
ACCESSION   AC112040
VERSION     AC112040.2  GI:21737402
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 119630)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joubah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisegeд,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Pacton,B., Payton,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rojas,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119630)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 119630)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18701992.
            -------------- Genome Center
            Center: Baylor College of Medicine
```

```
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center project name: BCM
            Center clone name: CH230-61E10
            -------------- Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 48405 bases at least Q40
            Consensus quality: 52337 bases at least Q30
            Consensus quality: 56201 bases at least Q20
            --------------
            NOTE: Estimated insert size may differ from sequence length
              (see http://www.bcm.tmc.edu/docs/Genbank_draft_data.html).
            NOTE: This is a 'working draft' sequence. It currently
            consists of 59 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
```

```
                1  1359: contig of 1359 bp in length
             1360  1459: gap of unknown length
             1460  2850: contig of 1391 bp in length
             2851  2950: gap of unknown length
             2951  4544: contig of 1594 bp in length
             4545  4644: gap of unknown length
             4645  5975: contig of 1331 bp in length
             5976  6075: gap of unknown length
             6076  7435: contig of 1360 bp in length
             7436  7535: gap of unknown length
             7536  8535: contig of 1000 bp in length
             8536  8635: gap of unknown length
             8636 10194: contig of 1559 bp in length
            10195 10294: gap of unknown length
            10295 11310: contig of 1016 bp in length
            11311 11410: gap of unknown length
            11411 12651: contig of 1241 bp in length
            12652 12751: gap of unknown length
            12752 14281: contig of 1530 bp in length
            14282 14381: gap of unknown length
            14382 15543: contig of 1162 bp in length
            15544 15643: gap of unknown length
            15644 17091: contig of 1448 bp in length
            17092 17191: gap of unknown length
            17192 18716: contig of 1525 bp in length
            18717 18816: gap of unknown length
            18817 20199: contig of 1383 bp in length
            20200 20299: gap of unknown length
            20300 21839: contig of 1540 bp in length
            21840 21939: gap of unknown length
            21940 23412: contig of 1473 bp in length
            23413 23512: gap of unknown length
            23513 24818: contig of 1306 bp in length
            24819 24918: gap of unknown length
            24919 26138: contig of 1220 bp in length
            26139 26238: gap of unknown length
            26239 27593: contig of 1355 bp in length
            27594 27693: gap of unknown length
            27694 29552: contig of 1859 bp in length
            29553 29652: gap of unknown length
            29653 31301: contig of 1649 bp in length
            31302 31401: gap of unknown length
            31402 32859: contig of 1458 bp in length
            32860 32959: gap of unknown length
            32960 34269: contig of 1310 bp in length
            34270 34369: gap of unknown length
            34370 35833: contig of 1464 bp in length
            35834 35933: gap of unknown length
            35934 37227: contig of 1294 bp in length
```

```
RESULT 14
AF068627

Qy   2599  CCTGGCATGAACAGGCCCTTTGGCCATCCACTGTGTGAATGATAAAGCTGGAGCTGCAAGAGTGT  2658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56266  ATGATCGACGCCAAAGAAGTGTCTGCTCCGACACACAGGGCCCGTTTACTTCTGGGGTAACCTT  56207

Qy   2539  ATGATTGACGCCAAAGAAGTGTCTGCTGCACACAGGGCCCGTTTACTTCTGGGGTAACCTT  2598
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56326  GTGGGCCATGGGCGTTAATGACAAGAGGGACATCTCCGCGATTTCTTCAGTGTCTAACCCCCGTG  56267

Qy   2479  GTGGGCCATGGGCGGTTAGTGACAAGAGGGACATCTCCGCGATTTCTTCAGTGTCTAACCCCCGTG  2538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Query Match   22.9%;   Score 962;   DB 2;   Length 119630;
Best Local Similarity 92.5%;   Pred. No. 2.1e-218;
Matches 1108; Conservative 0; Mismatches 60; Indels 30; Gaps 8;

```
*      37228    37327:  gap of unknown length
*      37328    38672:  contig of 1345 bp in length
*      38673    40409:  gap of unknown length
*      38772    40510:  contig of 1637 bp in length
*      40410    40509:  gap of unknown length
*      40510    42386:  contig of 1877 bp in length
*      42387    42486:  gap of unknown length
*      42487    43891:  contig of 1405 bp in length
*      43892    43991:  gap of unknown length
*      43992    45547:  contig of 1556 bp in length
*      45548    45647:  gap of unknown length
*      45648    47802:  contig of 2155 bp in length
*      47803    47902:  gap of unknown length
*      47903    49367:  contig of 1465 bp in length
*      49368    49467:  gap of unknown length
*      49468    51732:  contig of 2265 bp in length
*      51733    51832:  gap of unknown length
*      51833    53464:  contig of 1632 bp in length
*      53465    53564:  gap of unknown length
*      53565    55043:  contig of 1479 bp in length
*      55044    55143:  gap of unknown length
*      55144    55671:  contig of 1528 bp in length
*      55672    55771:  gap of unknown length
*      55772    57608:  contig of 1837 bp in length
*      58609    58708:  gap of unknown length
*      58709    60082:  contig of 1374 bp in length
*      60083    60182:  gap of unknown length
*      60183    61662:  contig of 1480 bp in length
*      61663    61762:  gap of unknown length
*      61763    64208:  contig of 2446 bp in length
*      64209    64308:  gap of unknown length
*      64309    66380:  contig of 2072 bp in length
*      66381    66480:  gap of unknown length
*      66481    69078:  contig of 2598 bp in length
*      69079    69178:  gap of unknown length
*      69179    71272:  contig of 2094 bp in length
*      71273    71372:  gap of unknown length
*      71373    73377:  contig of 2005 bp in length
*      73378    73477:  gap of unknown length
*      73478    75898:  contig of 2421 bp in length
*      75899    75998:  gap of unknown length
*      75999    77816:  contig of 1818 bp in length
*      77817    77916:  gap of unknown length
*      77917    80747:  contig of 2831 bp in length
*      80748    80847:  gap of unknown length
*      80848    82669:  contig of 1822 bp in length
*      82670    82769:  gap of unknown length
*      82770    84419:  contig of 1650 bp in length
*      84420    84519:  gap of unknown length
*      84520    87705:  contig of 3186 bp in length
*      87806    87805:  gap of unknown length
*      87806    90005:  contig of 2200 bp in length
*      90006    90105:  gap of unknown length
*      90106    92538:  contig of 2433 bp in length
*      92539    92638:  gap of unknown length
*      92639    95113:  contig of 2475 bp in length
*      95114    95213:  gap of unknown length
*      95214    98154:  contig of 2941 bp in length
```

```
Db  56206  CCTGGCATGAACAGGCCCATTGGGCATTCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGT  56147

Qy   2659  CTGGGAGCCACGGCAGAATAGCCAAGTTCAGCAAAAGTGAGGCACCATTACCACCAGGTCAAAC  2718
           |||||   ||||
Db  56146  CTGGAACACG--------GCCAAGTTCAGCCAAAACTGAGGCCATTACCACCAGGTCAAAC  56095

Qy   2719  TCTATAAAGCAGGGCAAAGACCAGCAATTTCCCGTCTTCATGAACGAGAAGAGGACATC  2778
           ||| ||||||||||||||||||||||||||||||||||| ||||||||||||
Db  56094  TCCATAAAGCAGGGCAAAAGACCAGCAATTTCCCGTCTTCAATGAATGAAGAGGACATC  56035

Qy   2779  CTGTGGTGCACTGAAAATGGAAAGGGTGTTTGGCTTCCCGTCCCACTCACACAGACGTCTCC  2838
           |||||||||||||||||||||||||||||||||||||||| ||  |||||
Db  56034  CTCTGGTGCACTGAAAATGGAAAGGGTGTTTGGCTTCCCGTCCCACTCACACAGACGTCTCC  55975

Qy   2839  AACATGAGCCCCGTTGGGCGGGAGGCGAGACGGGACTGCGTTGGGGGTCGCCGGTCATC  2898
           |||| ||  |||||||||||||||||||||||||||||||||| ||  ||||
Db  55974  AACATGACGCCCGTCGGGCAGGACACTTGCTAAGGCGGAACTTACCAGTCACC  55915

Qy   2899  CGCCACCTCTTCGCTCCGGCTGAAGGAAATATTTTGCTTGTCGTGTAAGGGACATGGGGGCAA  2958
           ||| ||||||||||||||||||||||||||||||||| |||||||||
Db  55914  CGCAACCTTTTCGCCGGGCTGAAGGAAATATTTTGCTTGTCGTGTAAGGGACATGGGGGCAA  55855

Qy   2959  ACTGAAGTAGTGATGAATGAATAAAAAGTTAAACAAAACAAAAACAAAAAACAAAAACA  3018
           |||||||||||||||||||||||||||||||||| ||||||    |||||
Db  55854  ACTGAAGTAGTGATGAATAACCAAGTTAAACAAAAACAAAA--------AACACA  55805

Qy   3019  ATAAAACACCAAGAAACGAGAGAGGAGGCGGAGAGAAAAGTTCAGCACCCAGAAGAGGAAT  3078
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55804  ATAAAAACACCAAGAAACGAGAGAGGAGGACGGGAAGAGTTCAGCACCCAGAAGAAAGGAAT  55745

Qy   3079  TTAAAGCAAAACACACAGAGGAGGGAGAAAACCGCGGAGGGCTTGGCCTTGCAAAAAGGTTGAC  3138
           ||| || ||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  55744  TTAATGCGAAACCACACAGAGGAGG--AAATGCCGGAGGGCTTGGCTGTCGCAAAAGGTTGGAC  55686

Qy   3139  ATCATCTCCTGAGTTTTCAAATGTTAAACCTTCAGTCCTATCTAAAAAAGCAAAAATAGGCCCC  3198
           ||||||||||| ||||| |||||||||||| |||||||||||||||  ||||||||||
Db  55685  ATCAATCTCTTGAGTTTTCAAATGTTCATCTTCAGTCCTAACTAACAAAGCAGGATAGGCCCC  55626

Qy   3199  TCCCCCTTCTTCCCCGTCCTCCGGTGCCTPAGGAGCGGAACTTTTTGTTTTTCTACTCTTTTCAGAG  3258
           ||||||||| ||||||||||||||||||||| |   |||
Db  55625  TCCCCCTTCCCCCCTCCGCGGTGTCCCGGAGGTGAACCCTTTTGTTTTTTCTACTCTTTTCAG  55566

Qy   3259  -GGGTTTTCGTTGTTGTTGGGGTTTTTTTTTTTTTGCTGTGAAACAAGAGAGAGTT-ATT  3316
           ||| |||||||| ||||||| |||||||||||||||||||||||||||||||||||||
Db  55565  AGGGGTTTCTGTTTTTTGTTGGGTTTTGTGTTTTTTTTTGCTGTGACTGAAACAAGAGAGTTAATT  55506

Qy   3317  GCAGCAAAAAATCAGTAACACAACAAAAAGTAGAAAAATGCCTTGGAGGAGGAAAGGAGGAGGGA  3376
           ||||| |||| ||| |||||||||||||||||||||||||||||||||||||||||||||
Db  55505  GCAGCAATATCAGTAACAACAAAAAAGTAGTAAATGGCCTTGGAGGAGGAAAGGAGGAGGGA  55446

Qy   3377  AAATTCTATAAAAACTTAAAAATATTGGTTTTTTTTTTTTTTCCCTTTTCTATATATCTCTT  3436
           ||||| ||| |||  |||||||||| ||||||||||||||||| ||||||||
Db  55445  AAATTCTATAAAAACTTAAAAAATAIITG--------TTTTTTTTCCCTTTTCTATATATCTCTT  55392

Qy   3437  TGGTTGTCTCTCTAGATCAGATGAATAGGAGCACAAAAACAGGAAGAAAATAGAGACCCTCGGA  3496
           ||||||||||||| |||||||||||||||||||||||||||||||||| ||||||||
Db  55391  TGGTTGTCTCTCTAGCCTGGTTCAGATAGGCAGCAACAAACAGGAAGAATAGAGACCTTCGGA  55332

Qy   3497  GGCAGAGTCTCCCTCTCCCACCCCGAGCAGTCTCAACAGCACCATTCCTGGTTCATGCAA  3556
           ||||||||||||||||| | |  |||||||||||||||||||||||||| |||||||||
Db  55331  GGCAGAGTCTCCTTCTTGCAGACCCGAGCAGTCTCAACAGCACTAGTTCCTGATCAGTGCAG  55272

Qy   3557  AACAGAAACCAACTAGCAGCGCGGCCGCTGAGAGAACAGACACCCACACAGACAC-TTTCTACAG  3615
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  55271  AACAGAACCCAACTAGCAGCAGCGGGCGCT--GAGAACGACCCACCAGACAACTTTTCTACAG  55214

Qy   3616  TATTTCAGGTGCCTACCACACAGGAAAACCTTGAAGAAAACACCAGTTTCTAGAAGCCGCT  3673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55213  TATTTCAGGTGCCTACCACACAGGAAAACCTTGAAGAAAACACCAGTTTCTAGAAGCCGCT  55156
```

```
LOCUS       AF068627                4135 bp    mRNA    linear   ROD 06-DEC-1999
DEFINITION  Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF068627
VERSION     AF068627.2  GI:6449471
KEYWORDS    .
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4135)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE     Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
  JOURNAL   Nat. Genet. 19 (3), 219-220 (1998)
  MEDLINE   98324766
   PUBMED   9662389
REFERENCE   2  (bases 1 to 4135)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REFERENCE   3  (bases 1 to 4135)
  AUTHORS   Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK      Sequence update by submitter
COMMENT     On Nov 18, 1999 this sequence version replaced gi:3327981.
FEATURES             Location/Qualifiers
     source          1..4135
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="84.0 cM"
     gene            1..4135
                     /gene="Dnmt3b"
     misc_feature    1..4135
                     /gene="Dnmt3b"
                     /note="similar to EST sequences deposited in GenBank
                     Accession Numbers AA116694, AA119979, AA177277, AA210568,
                     AA407106, and AA575617"
     CDS             269..2788
                     /gene="Dnmt3b"
                     /function="de novo DNA methylation"
                     /note="alternatively spliced product; contains Cys-rich
                     region; C-terminal region is similar to corresponding
                     region of DNA cytosine-5 methyltransferase 3A"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3B2"
                     /protein_id="AAC40179.2"
                     /db_xref="GI:6449472"
                     /translation="MKGDSRHLNEEEGASGYEECIIVNGNEFSDQSSDTKDAPSPVLE
                     AICTEPVCTPETRGRRSSSRLSKREVSSLLNYTQDMTGDGRDRDEVDDGNGSDILMPK
                     LTRETKDTRTRSESPAVRTRHSNGTSSLERQRASPRITRGRQGRRHVQEYPEVEFPATR
                     SRRRASSSASTPWSSSPASVDEMEEVTPKSVSTPSVDLSQDQDEGMDTTQVDAESRD
                     GDSTEYQDDKEGIGIDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                     SEISADKLVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTFPSSSPGESLED
                     QLKPMLEWAHGGFKPTGIEGLKPNKKQPENKSRRTTNDSAASESPPPKRLKTNSYGG
                     KDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVSFHPLFEGGLCQSCRDRFLEL
                     FYMYDEDGYOSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGAGTAEDAKLQEPWS
                     CYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPDLEEFEPPKLYPAIPAAKRRPIRVLSL
                     FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVWHEGQIKYVNDVRKITKKNIEE
                     WGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYRLLNYTRPKEGDNRPFFWMF
                     ENVVAMKVNDKDISRFLACNPVMIDAIKVSAAHRARYFWGNLPCMNRPVMASKNDKL
                     ELQDCLEERSRTAKLKKVQTITFKSNSTRQGKNQLFPVVMNGKDDVLMCTELERIGFP
                     AHTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDIFACE"
ORIGIN
BASE COUNT   1043 a   1048 c   1106 g    938 t
```

```
Qy      991  GAGCCCAGTAGGAGGGGGATGCTGGGGGACAAGAGAATGCTACCAAAAGCAGCCGGACGATGAGCCT 1050
             |||||  ||    ||   || |||||   |||||
Db      875  GATCAGGAGGGTATGGATACCACACAGCGTGGATGCAGAGAAGTGGGAGCAGAGATGGAGACAGACA  934

Qy     1051  GAGTATGAGGAATGGCCCGGGCGGGGCTTTGGCCATTGGAAGAGAGCTGGTTGTGGGGGAAACTTCGGGGC 1110
             ||||||  ||||||  ||||||| ||||||||||||||||||||||||||||||||||||||||| |||
Db      935  GAGTATCAGGATGATGAATAAAGAGTTTGAAATAGGTGACCTCGTGTGTGGGGAAAGATCAAGGGC  994

Qy     1111  TTCTCCTGGGTGGCCAGGCCGGAATTGTGTCTTGGTGGGATGACAGGCCGGAGCCTGGGAAGCCAGCT 1170
             |||||||||||||  ||||||||| ||||||||||||||||||||||||||||
Db      995  TTCTCCTGGGTGGCCTGCCATGGTGGTGTTCCTCGGAAAGCCACCTCCAAGCACAGGCCATG 1054

Qy     1171  GAAGGCACTCCGCTGGGTCATCGTGGTTCGGAGAGATGGCAAGGTTCTCAGTGGTGGTGTGGAG 1230
             ||     |||||||||||||     |||||||||  ||||| |||||||| ||||||||| |||
Db     1055  CCCGGAAATGCGCGTGGGTTACACGTGGTTTGGGTGATGCAAGGGTTTTTCTGACATCTTCTGCTGAC 1114

Qy     1231  AAGCTCATGCGCGTGAGCTCCTTCTGCCAGTGCATTCCACCAGGGCCACCTACAACAAGCAG 1290
             |||  |||||||  ||  ||||||||| ||||||||||||||||||
Db     1115  AAACTGGTGGCCTCTGGGGCTGTTCAGCCAGCACTTTAATCTGGCTTACCTTCAAATAAGCTG 1174

Qy     1291  CCCATGTACCCGCAAAAGCCATCTACGAAGTCCTCCAGGTGGCCCAGCAGCCGTGCGGGAAG 1350
             ||      ||||||||||||||      ||||||||||||||||||||||||||||||||| |||
Db     1175  GTTTCTTATAGGAAAGGCCCATGTACCACACTCTGGAGAAACGCAGGGGTTCGAGCGTGGCCAAG 1234

Qy     1351  CTGTTTCCCAGCTTGCCATGACAGTGAATGAAAAGTGACAGTGGTGCAAGGCTGTGGAAGTGAAG 1410
             ||||||||||||||||  ||||||||||||
Db     1235  ACCTTCTC---------CAGCAGTCCTGGAGAGTCACTGGAGGACCAG 1273

Qy     1411  AACAAGCAGGATGCATTGAATGGGCCCTCGGTTGGCTTCCAGCCCTCGGGTTCCTAAGGGCCTG 1470
             ||    ||||||||||||||||||||||||| |||||||||||||||||||||||
Db     1274  CTGAAAGCCCATCCTGGAGTGGGGGCCCACGGGTGGCTCAAAGCCTACTGGGATCGAGGGCCTC 1333

Qy     1471  GAGAACACCAGAGAAGAGAGAGAAGAAATCCTTTACAAAGAGAAGTTTACCACGACAATGTGGGTGACA 1530
             ||    |||||||||||||||    |||||| || ||||
Db     1334  AAACC------CAACAAGAAGCCAACCCAGAGAAACAAAAGTCGAAGAACGACACAACCAATGAC 1387

Qy     1531  CCTGAAGCAGCTGCTTACGCCCCACCCCCAACCCAGCCCAGAAAACCCAGAAAGAGACACAACA 1590
             ||| ||     ||  |||  ||||||||||||||||||||||||||
Db     1388  TCTGCTGCTTCCTGGAGTCCCCCCCACCCCCAACGCCCTTCGTCGACAAATAGCTATTGCGGGAAG 1447

Qy     1591  GAGAAACCTAAGGTCAAAGGAGATCATTGATGAGCCCACAAGGGAGCGGCTGTTGTATGAG 1650
             ||  |||| ||||| ||||||  ||||||| ||||
Db     1493  GTCACCAACACAACAAGGGCAATCTGAAGAACCGCGTGTTTGTCCGTGTGGAAAGAGAACCCT 1552

Qy     1651  GTGCGGCCAGAAGTGCAGAAAACATCGAGGACATTTGTATCTCCATGTGGGAGACCTCAATGTC 1710
             || ||||||||||||||||||||||||||||||||||||||||| ||||
Db     1493  GTCACCAACACAACAAGGGCAATCTGAAGAACCGCGTGTTTGTCCGTGTGGAAAGAGAACCCT 1552

Qy     1711  ACCCTGGAGCACCCCACTCTCTTCATTGGAGGCATGTGCCCAGAACTGTAAGAACTGCTTCTTG 1770
             ||||||| |||||||||||||||||||||||||||||||||||||
Db     1553  GTGTCCTTCACACCCCTCTTTGAGGGTGGGCCTCTGTCACAGAGTTGCCGGGATCCGTCTGTTCCTA 1612

Qy     1771  GAGTGTGCTTACCAGTATGACCACGATGGGTACCAGTCCTATTGCACCATCTGCTGTGGG 1830
             ||||| ||||||  ||||||||||||
Db     1673  GCCCGTGAAACTGCTTGCTGTCTGTGGGAACACAACTGCTGCAGGTGCTTTTGTCGAGTGT 1732

Qy     1831  GGGCGTGAAGTGCTCATTGTGTGGGAACACAACTGCTGCAGGTGCTTTTGTCGAGTGT 1890
             || |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     1673  GCCCGTGAAACTGCTTGCTGTCTGTGCAGTAACACAAAGCTGCTGCAGATCGTCTTCTGTTGGTGGAGTGT 1732

Qy     1891  GTGGATCTCTTTGGTGGGGGCCAGGAGCTGCTTCAGGCCAGCCCATTAAGGAAGAAGCACCCCTGGAAC 1950
             |||  ||||||||||||||||||||||||||||||||||||
Db     1733  CTGGAGGTGCTGGTGGGCGCCAGGCACAGGTGAGCAGGTGTGACGACCCTGGAGC 1792

Qy     1951  TGCTACATGTGCGGCCATAAGGGCACCTATGGGCTTGCTCGGAACACGGGAAGGACTGGCCT 2010
             |||||||||||||||   || |||||||||||||||||||||||||||||||||||| |||
Db     1793  TGCTATATGTGCCTCCCTCAGCGCTGCCATGGGGTCCTCCTGCCGACCAGACGAAAAGATTTGGAAC 1852

Qy     2011  TCTCGAGTCAGATGTTCTTTGCCAATAACCATGACC---AGGAATTTGACCCCCCAAAG 2067
             |||||  ||||||||||||||||||||||||||||    ||||||||||||||  |||
Db     1853  ATGCGCCTGCAAGACTTCTTCACTACTGATCCTGAAGAAATTTGAGCCACCCAAG 1912
```

Alignment (left block)

```
Qy  2068  GTTTTACCCACCTGTGCCCAGCTGAGGAGGAGGAAGCCATCCGCGTGCTGTCTCTTTGAT  2127
Db  1913  TTGTGACCCAGCAATTCCTGCAGCCAAAAAGGAGGCCCATTAGAGTGCTGTCTCTGTTTGAT  1972

Qy  2128  GGGAATTGCTACAGGGCCTCCTGGTGCTGCTGAAAGGACCTGGGCGACCGCTACATT  2187
Db  1973  GGAATTGCAACGGGGTACTTGGTGCTCAAAGGAGTGGGTATTAAAGTGGAAAAGTACATT  2032

Qy  2188  GCCTCCGAGGTGTGTGAGGAGACTCCATCACGGTGGGCATGGTGCCGGCACCAGGGAAAGATC  2247
Db  2033  GCCTCCGAAGTCTGTGCAGAGTCCATCGCTGTGGGAACTGTTAAGCATGAAGGCCAGATC  2092

Qy  2248  ATGTACGTCGGGGACGTCCGCAGCGTCACACAGAAGCATATCCAAGGAGTGGGCCCATTC  2307
Db  2093  AAATATGTCAAATGACGTCCGGAAAAATCACCAAGAAAAATATTGAAGAGTGGGCCCGTTC  2152

Qy  2308  GACCTGGTGATTGGAGGGCAGTCCCTGCCAAATGACCTCTCCATTGTCAACCCTGCCCGGAAG  2367
Db  2153  GACTTGGTGATTGGTGGAAGCCCATGCAATGACCTCTCTAAACGTCAAATCCTGCCCGCAAA  2212

Qy  2368  GGACTTTATGAGGGGTACTGGGCCGCCCTCTTCTTTTGAGTTCTTACCGCCCTCTGCCATGATCGCG  2427
Db  2213  GGTTTATATGAGGGGCACAGAGGAAGGGCTCTTCTTTCGAGTTTTTACCACTTGCTGAATTATACC  2272

Qy  2428  CGGCCCCAAGGAGGGAGGAGGAGATGATCGCCCCTTCTTCTGGCCTCTTTGAGAATGTGGTGGGCCATG  2487
Db  2273  CGCCCCAAGGAGGGGGGACAACCGGTCCATTCTTCTTCTGGAGTTCCGAGAATGTTGTGGGCCATG  2332

Qy  2488  GGCGTTAGTCACAAGAGGGGACACTCTCCGGATTTTCTTGACTCTTAACCCCGTCATGATTGATTGAC  2547
Db  2333  AAAGTGAATGACAACGAAAGAGACAATCTCAAGATTCCTGGCAATTCCAGTAACCCCGGTCGAT  2392

Qy  2548  GCCAAAGAAGTGTCTCGCACACAGGGGCCCGTTACTTCTTCTGTGGGGTAACCTTCCTGGCCATG  2607
Db  2393  GCCATCAAGGTGTCTGCTGCACAGCGCCGGTACTTCTGGGTAACCTACCCGGAATG  2452

Qy  2608  AACAGGCCTTTGGCATCCACTGTGAATGATAAGCTGGACGAGAGTGTTCTGGAGCAC  2667
Db  2453  AACAGGCCCGTGATGGCTTCAAAGAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTC  2512

Qy  2668  GGCAGAAATAGCCAAGTTCAGCAAAGTGAGGACCATTACCACCAGTTCAAAACTCTATAAAG  2727
Db  2513  AGTAGGACAGCAAAGTTAAAGAAAGTGCAGACAAATAACCACCAAGTCGAACTCCATCAGA  2572

Qy  2728  CAGGGCAAAGAACCAGCATTTCCCGTCTTCATGAACGAGAAGGAGGACACATCCTGTGGTGC  2787
Db  2573  CAGGGCAAAAAACCAGCTTTTCCCTGTAGTCATGAAATGGCAAAGGACGACGTTTTGTGGTGC  2632

Qy  2788  ACTGAAAATGGAAAAGGGTGTTTGGCCTTCCCCGTCCACTACACAGACGACGCTCTCCAACATGAGC  2847
Db  2633  ACTGAGGCTCGAAAAGGATCTTCGGCTTCCTGCTCACTACACGGACGTGTCCAACATGGGC  2692

Qy  2848  CGCTTGGCCAGGAGGACGACTTCTCTGGGGCCGATCTGTGGGACGTGCCCGGTCATCCGCCACCTC  2907
Db  2693  CGGCGGCGCCCGTCAGAAGCTGTCGGAGTGTACCGGTCATCAGACACACCTG  2752

Qy  2908  TTCCGCCTCCGCTGAAGGAAATATTTGCTTGTGTTA  2942
Db  2753  TTTGCCCCCTTGAAGGACTACTTCGCTGTGAATA  2787
```

GenBank record

```
REFERENCE   1  (bases 1 to 4163)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE     Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4163)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China

FEATURES             Location/Qualifiers
     source          1..4163
                     /organism="Mus musculus"
                     /strain="KM"
                     /db_xref="taxon:10090"
                     /dev_stage="8-9 day old embryo"
     gene            1..4163
                     /gene="Dnmt3b"
     CDS             297..2816
                     /gene="Dnmt3b"
                     /function="methylates cytosine in DNA"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA cytosine-specific methyltransferase isoform
                     6"
                     /protein_id="AAF74520.1"
                     /db_xref="GI:8347131"
                     /translation="MKGDSRHLNEEGSASGYEECIIVNGNFSDQSSDSDTKDAPSPPVLE
                     AICTEPVCTPETRGRSRSSSRLSKREVSSLLNYTQDMTGDGDRDDEVDDGNGSDILMPK
                     ITRETKDTRTRSSEPAVRTRHSNGTSSTSSLERQRASPRITRGRQGRHHVQEYPVEFPATR
                     SRRRASSSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDEGMDTQVDAESIY
                     GDSTEYQDDKEFGIGDLVWGKIKGFSWNPAMVVSWKATSKROAMPGMRWVQWFGDGFF
                     SEISADKIVALGLFSQHFNLATFENKLVSYRKAMYHTLEKARVRAGKTFSSSPGESLED
                     QLKPMLEWAHGGFKPTGIEGLKPNKQPENKSRRRTTNDSAASESPPKRLKTNSYGG
                     KDRGEDEESRERMASEVTNNRGNLEDRCLSCGKKNPVSPHPLFEGGLCQSCRDFLEL
                     FYMYDEDGYQSYCTVCCEGRELLLCNTSSCCRCPCVECLEVLVGAGTAEDAKLQEEPWS
                     CYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPDLEEFEPKLIYPAIPAAKRRPIRVLSL
                     FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYNDVRKTITKKNIEE
                     WGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEYYHLLNYTRPKEGDNRPFFWMF
                     ENVAMKVNDKDISRPLACNPVMIDAIKVSAAHRARYFWGNLPGMNRPVMASKNDKL
                     ELQDCLEFSRTAKLKKVQTIITTKSNSIROGKNQLFPVVWNGKDDYLWCTELERIFGFP
                     AHYTDVSNMGRGAROKLLGRSWSVPVIRHLFAPLKDYFACE"

BASE COUNT     1032 a   1074 c   1128 t    929 g
ORIGIN
```

RESULT 15

```
AF151974

LOCUS       AF151974                4163 bp    mRNA     linear   ROD 08-JUN-2000
DEFINITION  Mus musculus DNA cytosine-specific methyltransferase isoform 6
            (Dnmt3b) mRNA, complete cds.
ACCESSION   AF151974
VERSION     AF151974.1  GI:8347130
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

Query Match 19.0%; Score 794.6; DB 10; Length 4163;
Best Local Similarity 64.8%; Pred. No. 1.5e-178;
Matches 1266; Conservative 0; Mismatches 644; Indels 45; Gaps 4;

Alignment (second block)

```
Qy   991  GAGCCAGTAGGAGGGGAATGCTGGGGACAAGAAATGCTACCAAAGCAGCCGACGGATGAGCCT  1050
Db   903  GATCAGGAGGGGTATGGATACCACACAGGTGGGATCCAGAGAGCAGCATATATGGAGACAGCACA  962

Qy  1051  GAGTATGAGGAGATGGCGCCGGGGCTTTGACGACGCTGCTGTGGGGGAAACTTCGGGGC  1110
Db   963  GAGTATCAGGATGATGAAAGAGAGTTTTGGAATAGGTGAACCTTCGTGTGGGGAAGAATCAAGGGC  1022

Qy  1111  TTCTCCTGGGTGGCCAGGCCGGAAGTGTGTCTTGGTGGATGGACAGGCCGGAGCCAGCT  1170
Db  1023  TTCTCCTGGTGGCCTGGCCACTGGGTGGGTGTCCTGGAAAGCCACCTCCAAAGCCGGAGCCCATG  1082

Qy  1171  GAAGGCACTCGCCTGGCTGGTGGTTCGGAGAGTGGCAAACTTCTCAGTGGTGTTGTGGAG  1230
Db  1083  CCCGGAAATGCCCTGGGTGGGTACAGTGGTGTTTTGGTGATGGCACAAGTTTTTCTGAGATCTGCTGAC  1142

Qy  1231  AAGCTCATGCCGCGCTGAGCCTCCTTTCTGCAGTGCATTCCACCAGGCCACCTACACAGGCAG  1290
Db  1143  AAACTGGTGCGCCTCGGGGGCTGTTCAGCCAGGCACTTTAATCTGGCTTCAAATAAGCTG  1202
```

```
Qy   1291 CCCATGTACCGCAAAGCCATCTACGAAGTCCTCCAGGTGGCCAGCCAGCCCGTGCCGGGAAG 1350
               ||||||   ||| |  |||    |  |  || |||  |||    |||   |||  || ||
Db   1203 GTTTCTTATAGGAAGGCCATGTACCACACTCTGGAGAAAGGCAGGGTTCCAGCTGGCAAG 1262

Qy   1351 CTGTTTCCAGCTTGCCATGACAGTGATGAAAGTGACAGTGCAAGGCTGTGGAAGTGCAG 1410
               |||                         ||                  |||
Db   1263 ACCTTCTC------------------------CAGCAGTCGTGGAGAGTCACTGGAGGACCAG 1301

Qy   1411 AACAAAGCAGATGATTGAATGGGGCCCTTCCAGCCCTTCGGGTCCTTAAGGGCTG 1470
               |||   ||  || ||  ||||       | |    ||        ||   |||
Db   1302 CTGAAGCCCATGCTGGAGTGGGCCCACGGTGGCTTCAAGGCTCACCTACTGGAGGGCCTC 1361

Qy   1471 GAGCCACCAGAAGAAGAGAAGAGAAATCCTTACAAAGGAAGTTTACACCGGACATGTGGGTGGAG 1530
               ||| ||   | |  |    ||  |||    |||  || ||  |||  |||||| ||
Db   1362 AAACC------CAACAAGAAGCAAACCCGAGAACAAAAGTCGAAGACACCACAACCAAATGAC 1415

Qy   1531 CCTGAAGCAGGCTGCTTACGGCCCCCACCCCCACCCAGCCCAAGAAACCCAGAAAAGAGACAACA 1590
               |||      |||  ||| |    || ||| |||| ||  ||||  |||| ||||| ||
Db   1416 TCTGCTGCTTCTTGAGTCCCCCCACCCACCCAAGCGCCTCAAGACAAAATAGCTATGGCGGGAAG 1475

Qy   1591 GAGAAACCTAAGGTCAAGGAGAGATCATTGATGAGCCGACACAAGGACGCGGTGGTTGGTGTGAG 1650
               ||| ||| |||   |||  || |||  || | |    || ||  ||| |   | |||
Db   1476 GACCGAGGGGGAGGGATGAGGGAGC-------------------CGAGAACGGATGGCTTCTGAA 1520

Qy   1651 GTGCGCCAGAAGTGCAGAAACATCGAGGACACATTTGTATCTCATGTGGGAGCCTCAAATGTC 1710
               |||  ||   ||  |     ||||   || |||    | || |  || |  |||  |||
Db   1521 GTCACCAACAACAAGGGCAAATCTGGAGAACCGCTGTTTTGTCCTGTGGAAAGAAGAAACCCT 1580

Qy   1711 ACCCTGGAGCACCACTCTTCATTGGAGGCATGTCCAGAACTGTAAGAAACTGGTTCTTG 1770
               |||| || |  ||  |  ||  ||  |   | |   || |   || |||  |||
Db   1581 GTGTCCTTTCCACCCCCTCTTTGAGGGTGGGCCTCTGTCAGAGTTGCCGGGATCGGCTTCCTA 1640

Qy   1771 GAGTGTGCTTACCAGTATGACGACGATGGGTACCAGTCCTATTGCCACCATCTGCTGTGTGGG 1830
               ||  ||  ||  |||  ||  ||   |||| |||| || |  || |||   ||| ||
Db   1641 GAGCTCTTCTACATGTATGATGAGGACGGGCTATCAGTCCTACTGCAGCCGGTGCTTGTGAG 1700

Qy   1831 GGGCGTGGAAGTGCTCATGGTGTGGGAACAACAACTGCTGCAAGGTGGCTTGTTTGTGCGAGTGT 1890
               ||| |  |||||||  |  ||||| ||  ||    ||| ||  || | ||| || |||
Db   1701 GGCCGTGAAACTGCTCTGCTGTGTGCAGATGCTCTTCTGTGTGGGAGTGT 1760

Qy   1891 GTGGGATCTCTTGGGTCGGGGCCACAGGAGCTGCTCAGGCAGGCCATTAAGGAGAAGACCCCTGGGAAC 1950
               ||||||   || ||| ||  || ||  | ||||   | |   ||     || |||     |||
Db   1761 CTGGAGGTGCTGGTGGGCGCAGGCACACTGACGTAGGCCAAGCTCCAGGAAACCCTGGAGC 1820

Qy   1951 TGCTACAATGTCGCGGGCATAAGGGCCACTTATGGGGCTGCTGGGAAGACGGGAAGACTGGCCT 2010
               |||   ||| ||||     ||  || || |  |  |   ||  ||   |  |||| ||||
Db   1821 TGCTATATGTGCCTCCCCTCAGCGCTCCCATGGGGTTCCTCCGACGCCGAGAAAGATTGGAAC 1880

Qy   2011 TCTCGACTCCACGATAGTTCTTTTGCCAATAACCATGACC---AGGAATTTGACCCCCCAAAG 2067
               |||  ||    |||  ||  ||   ||  | ||        |||||||  | |||
Db   1881 ATGGCGCCTGCAAGACTTCTTCACTACTGAATCCTGACCTGGAAGAATTTGAGCCACCCAAG 1940

Qy   2068 GTTTACCCACCTGTGCCAGCTGAGAAGAGACCATCCGCGTGCTGTCTGTCTGTTTGAT 2127
               || ||||  || | ||  ||| | |  |  ||||  | ||  |||   |  |||||
Db   1941 TTGTACCCAGCAATTCCTGCAGCCAAAAAGAGGAGGCCCCATTAGCAGTCCGTCTTGTGAT 2000

Qy   2128 GGGATTGCTACAGGGGCTCCTGGTGTGTGAAGGAGACCTGGGGGCCCCATTC 2187
               |||  ||| |  |    ||     | || ||| ||  | ||||  ||||||
Db   2001 GGAATTGCAACGGGGTTACTTGGTGCTCAAGGAGGAGAACTGGTTATTAAAAGTGGAAAGTACATTI 2060

Qy   2188 GCCTCCCAGGGTGTGTGAGGACTCCATCACGGTGGGCGCATGTGTGCGGCACCAGGGGAAAGATC 2247
               ||||||  || || ||||| |  |||| ||  |||||  || || ||||   |||
Db   2061 GCCTCCCGAAGTCTGTGGGGCAGTCCATCGCGTGTGGGGAACTGTTTAAGCCATGAAAGGCCAGATC 2120

Qy   2248 ATGTACGTCGGGGGACCTCCCGCAGCCGTTCACACAGAGAGCATATCCAGGAGTGGGGCGCCCCATTC 2307
               |||||||  |  ||   |  ||  ||  |   |   |  ||| | ||  ||||
Db   2121 AAATAGTCAATGACAGTCCCGGAAAAATCACCAAGAGAAAAATATTGAAGAGTGGGGCCCCGTTC 2180

Qy   2308 GACCTGGTGATTGGAGGCAGTCCCTGCCAATGACCTCTCCAATTGTCAACCCTGCCCGCAAG 2367
               ||||||| |  ||  ||| |  |     ||      |    |    |||| ||||  |
Db   2181 GACTTGGTGATTGGTGGAAGCCCATGCAAATGAATCTCTCTAACGTCAAATCCTGCCCGCAAA 2240

Qy   2368 GGACTTTATGAGGGTACTGGGCCGCCTCTTCTTTTGAGTTCTACCCGGCCTCCTGCCATGAATGCG 2427
```

```
Db   2241 GTTTTATAGGGGCACAGGAAAGGGCTCTTCTTCTTCGAGTTTTTACCACTTGCTGAATTATACC 2300
               |   ||||||||| | ||||||  || |  | |  |||||    ||| || |  |||||
Qy   2428 CGGCCCAAGGAGGGGAGAGATGGATCGCCCCCCTTCTTCTTCTCGAGTTTTGAGAATGTGGGTGGCCATG 2487
               ||||  ||||| ||||||| | ||  ||  ||| |||  ||||  | |||||  ||||
Db   2301 CGCCCAAGGGAGGGCGACAACCGTCCATTCTTCTTCTGATTGTTCGAGAATGTTGTGGCCATG 2360
               ||| ||||  |||  || ||  || |   ||  || | |  ||||
Qy   2488 GGCGTTAGTGACAAGAGGGACAATCTCCCGATTTCTTGAGTCTAACCCCGTGATGATTGAC 2547
               || || ||||  | |  |||  |||||   || ||| |  ||||   | |
Db   2361 AAAGTGAATGACAAGAGAAGACATCTCAAGATTCCTGGCATGTAAACCCAGTGATCATCGAT 2420
               ||| || || | | |||  | ||| ||   || ||  |  |  ||
Qy   2548 GCCAAAGAAGTGTCTGCTGCCACACAGGGCCCCGTTACTTCTGGGGTTAACCTTCCTGGCCATG 2607
               |||  ||||||  |||  ||||  |||   || | |  |  || |
Db   2421 GCCATCAAGGTGTTCTGCTCTCACAGGGCCCGTACTTCTGGGGTTAACCTTACCCGGAAATG 2480
               ||| ||||||   |||  | |  |  |  |  |||  |  ||||
Qy   2608 AACAGGGCCTTTGGCATCCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCAC 2667
               ||||     ||  |  |||| |  ||||| || ||| |  |||   |||||
Db   2481 AACAGGCCCGTGATGGCTTCAAAGAATGATAAAGCTCGAGCTGCAGGACCTGCCTGAGTTC 2540
               ||||| |  |   |  |   ||||  | |||  |||| ||||
Qy   2668 GGCAGAATAGCCAAGTTCAGCAAAGTGAGGGACCATTACCACCAGGTCAAACTCTATAAAG 2727
               |||||    ||  ||||| |  | || | |  ||||| || ||  |  |||
Db   2541 AGTAGGCACAGCAAAGTTAAAGAAAGTCAGACAAATAACCACCACCAACTGCACTCCATCAGA 2600
               |||  ||  || ||||| | ||  |  | ||  ||| ||||
Qy   2728 CAGGGCAAAGACCAGCACATTTCCCCGTCTTCAATGACGAGAAGAAGGAGGGACAATCCTGTGGGTGC 2787
               ||||| || |  ||||  ||  | ||| |||| || |  |||||| |||| ||
Db   2601 CAGGGCAAAAACCAGCTTTTTCCCTGTAGTCATGAAATGGCAAAGGACGAGCTGTTTTGTGGGTGC 2660
               ||||| |  |  ||||  || || |||   ||  | | ||||  || |
Qy   2788 ACTGAAAATGGAAAAGGGTTGTTTTGGCTTCCCCGTCCCACTACACACAGACAGTCTCCAACATGAGC 2847
               |||| |  ||| | |  |  ||  ||  || |  || |||   |  || | |||
Db   2661 ACTGAGCTCGAAAGGATCTTCGGGTTCCCTGCTCACTACACAGCAGCGTCCAACATGGGC 2720
               |||| |  || ||  | ||  ||||| || || |||||   |||||
Qy   2848 CGCTTGGGAGGCGAGAGACTGCTGTGGGCGGCCGATCGTGTGGAGGGTCATCCGCCACCTC 2907
               |||  | ||||  |||||| |  |||||| || ||  | || ||| ||
Db   2721 CGCGGCGCCCGCGTCAGAAGCCTGCGGGGCAGGTGTCCTGGAGTGTACCGGTCATCAGACACACCTG 2780
               ||| | |  |||  | ||| ||| |   |||| | |  || ||
Qy   2908 TTCGCTCCGCTGAAGGAAATATTTTTGCTTGTGTGTA 2942
               ||  | ||||  |  ||||| | ||||||| |  |||
Db   2781 TTTGCCCCCTTGAAGGACTACTTTGCCTGTGAATA 2815
```

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 22:42:04 ; Search time 7013.06 Seconds

                                   (without alignments)
                                   17408.415 Million cell updates/sec

Title: US-09-720-086-2
Perfect score: 4195
Sequence: 1 gaattccgggcgcgggggtt.........aaaaaaaaaaaaaaaaa 4195

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455140 2878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :

                GenEmbl:*
                1: gb_ba:*
                2: gb_htg:*
                3: gb_in:*
                4: gb_om:*
                5: gb_ov:*
                6: gb_pat:*
                7: gb_ph:*
                8: gb_pl:*
                9: gb_pr:*
                10: gb_ro:*
                11: gb_sts:*
                12: gb_sy:*
                13: gb_un:*
                14: gb_vi:*
                15: em_ba:*
                16: em_fun:*
                17: em_hum:*
                18: em_in:*
                19: em_mu:*
                20: em_om:*
                21: em_or:*
                22: em_ov:*
                23: em_pat:*
                24: em_ph:*
                25: em_pl:*
                26: em_ro:*
                27: em_sts:*
                28: em_un:*
                29: em_vi:*
                30: em_htg_hum:*
                31: em_htg_inv:*
                32: em_htg_other:*
                33: em_htg_mus:*
                34: em_htg_pln:*
                35: em_htg_rod:*
                36: em_htg_mam:*
                37: em_htg_vrt:*
                38: em_sy:*
                39: em_htgo_hum:*
                40: em_htgo_mus:*
                41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4195	100.0	4195	10	AF068626	AF068626 Mus muscu
2	4182.8	99.7	4338	10	AF151969	AF151969 Mus muscu
3	4065	96.9	4135	10	AF068627	AF068627 Mus muscu
4	4052.8	96.6	4278	10	AF151970	AF151970 Mus muscu
5	3942.8	94.0	4223	10	AF151973	AF151973 Mus muscu
6	3812.8	90.9	4163	10	AF151974	AF151974 Mus muscu
7	3807	90.8	4006	10	AY078427	AY078427 Mus muscu
8	3794.8	90.5	4149	10	AF151972	AF151972 Mus muscu
9	3677	87.7	3946	10	AF068628	AF068628 Mus muscu
10	3664.8	87.4	4089	10	AF151971	AF151971 Mus muscu
11	3554.8	84.7	4034	10	AF151975	AF151975 Mus muscu
12	3424.8	81.6	3974	10	AF151976	AF151976 Mus muscu
13	1815.4	43.3	4335	9	AF331857	AF331857 Homo sapi
14	1812.2	43.2	4145	9	AF156488	AF156488 Homo sapi
15	1697.2	40.5	4267	9	AF176228	AF176228 Homo sapi
16	1459.8	34.8	244329	2	AC107644	AC107644 Homo sapi
17	1335.8	31.8	3897	9	AF156487	AF156487 Homo sapi
18	1072.6	25.6	3017	9	AK001191	AK001191 Homo sapi
19	858.4	20.5	167568	2	AC111734	AC111734 Rattus no
20	762	18.2	3005	9	AF067972	AF067972 Homo sapi
21	762	18.2	4258	9	AF331856	AF331856 Homo sapi
22	761	18.1	4094	10	BC007466	BC007466 Mus muscu
23	759.4	18.1	4192	10	AF068625	AF068625 Mus muscu
24	696	16.6	2191	9	AK025230	AK025230 Homo sapi
25	656.2	15.6	2127	9	AF129267	AF129267 Homo sapi
26	548.4	13.1	2077	6	AR129189	AR129189 Sequence
C 27	537.8	12.8	176697	2	AC112586	AC112586 Rattus no
28	525.2	12.5	2848	5	AF135438	AF135438 Danio rer
29	520.6	12.4	2057	9	AF122268	AF122268 Homo sapi
30	516.2	12.3	1758	10	BC018214	BC018214 Homo sapi
C 31	435.6	10.4	188936	2	AC128509	AC128509 Rattus no
C 32	422.6	10.1	2008	9	AF122269	AF122269 Homo sapi
33	411.4	9.8	168651	2	AC009474	AC009474 Homo sapi
34	402.4	9.6	123936	2	AC120824	AC120824 Rattus no
C 35	266	6.3	119630	2	AC112040	AC112040 Rattus no
36	248.2	5.9	110000	9	AL354832_0	AL354832 Homo sapi
37	248.2	5.9	118899	9	HS1085E17	AL035071 Human DNA
38	196.4	4.7	49478	2	AC091678	AC091678 Mus muscu
39	196.4	4.7	244329	2	AC107644	AC107644 Mus muscu
C 40	191.8	4.6	49478	2	AC091678	AC091678 Mus muscu
C 41	191.8	4.6	225045	2	AL833803	AL833803 Mus muscu
C 42	188.6	4.5	249245	2	AC122356	AC122356 Mus muscu
43	187	4.5	225045	2	AL833803	AL833803 Mus muscu
44	181	4.3	210269	2	AC116459	AC116459 Mus muscu
C 44	178.8	4.3	139581	2	AC131350	AC131350 Rattus no
C 45	178.8	4.3	153533	2	AC095199	AC095199 Rattus no

ALIGNMENTS

RESULT 1
AF068626
LOCUS   AF068626   4195 bp   mRNA   linear   ROD 06-DEC-1999
DEFINITION   Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,
               alternatively spliced, complete cds.
ACCESSION   AF068626
VERSION   AF068626.2   GI:6449469
KEYWORDS   .
SOURCE   Mus musculus.
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4195)
AUTHORS   Okano,M., Xie,S. and Li,E.
TITLE   Cloning and characterization of a family of novel mammalian DNA

```
JOURNAL     (cytosine-5) methyltransferases
MEDLINE     Nat. Genet. 19 (3), 219-220 (1998)
PUBMED      98324766
REFERENCE   9662389
            2 (bases 1 to 4195)
AUTHORS     Xie,S., Okano,M. and Li,E.
TITLE       Direct Submission
JOURNAL     Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REFERENCE   3 (bases 1 to 4195)
AUTHORS     Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK      Sequence update by submitter
COMMENT     On Nov 18, 1999 this sequence version replaced gi:3327979.
FEATURES             Location/Qualifiers
     source          1..4195
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="84.0 cM"
     gene            1..4195
                     /gene="Dnmt3b"
     misc_feature    1..4195
                     /gene="Dnmt3b"
     CDS             269..2848
                     /gene="Dnmt3b"
                     /function="de novo DNA methylation"
                     /note="alternatively spliced product; contains Cys-rich
                     region; C-terminal region is similar to corresponding
                     region of DNA cytosine-5 methyltransferase 3A"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3B1"
                     /note="similar to EST sequences deposited in GenBank
                     Accession Numbers AA116694, AA119979, AA177277, AA210568,
                     AA407106, and AA575617"
                     /protein_id="AAC40178.2"
                     /db_xref="GI:6449470"
                     /translation="MKGDDRHLNEEEGASGYEECIIVNGNFSDQSSDTKDAPSPVLE
                     AICTEPVCTPETGRVRSRSSSRLSKREVSSLLNTYQDMTGDGDRDDEDVDDGNGSDILMPK
                     LTRETKDTRTRSSPAVRTRHSNGTSSLEROASPRITRGRQGRHHVQEYPVEFPATR
                     SRRRRASSSSASTPWSSPASVDFMEEEVTPKSVSTPSVDLSQDGDQEGMDTTQVDAESRD
                     GDSTEYQDDKEFGIGDLLWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                     SEISADKLVALGLPFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTFSSSPGESLED
                     OLKPMLEWAHGGFKPTGIEGLKPNKKQDPVVNKSKVRRSDSRNLEPRRENKSRRRTTN
                     DSAASESPPPKRLKTNSYGKDKGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVS
                     FHPLFEGGLCQSCRDRFLELFYMYDEDGYQSYCTVCCEGRELLLCSNSCCRCFCVBC
                     LEVLVGAGTAEDARLQEEPWSCYMCLPQRCHGVLRRRDWNMRLQDFFTTDDLEEFEP
                     PKLPAIPAAKRRPIRVLSLFDGIATGYIVLKELGIKVEKYIASEVCAESIAVGTVKH
                     EGOIKYVNDVRKITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFY
                     HLLNYTRPKEGDNPFFWMFENVVAMKVNDKDISRFLACNPVMIDAIKVSAAHRARY
                     FWGNLPGMNRPVMASKNDKLELQDCLEFSRTARLKKVQTITTKSNSIRQGKNQLFPVV
                     MNGKDDVLWCTELERIFGFPAHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYF
                     ACE"
BASE COUNT     1061 a   1059 c   1125 g    950 t
ORIGIN
```

Qy   Query Match
Best Local Similarity   100.0%;
Similarity   100.0%;   Pred. No.   0;
Matches 4195; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Score 4195;   DB 10;   Length 4195;

```
Qy     1 GAATTCCGGCGCGCGGGGGGGTTAAGCGGGCCCAAGTAAACGTAGCGCCAGCGAGCGCCGGA    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GAATTCCGGCGCGCGGGGGGGTTAAGCGGGCCCAAGTAAACGTAGCGCCAGCGAGCGCCGGA    60

Qy    61 GATTCGCGAACCCGACACTCCGCGCGCCGCGCCCGGCGCGGGCGCGGGGCGGGATCGCGGGC   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GATTCGCGAACCCGACACTCCGCGCGCCGCGCCCGGCGCGGGCGCGGGGCGGGATCGCGGGC   120

Qy   121 GCCGGCGCTACAGCCAGCCTCACGACGACGGGCGCCGCGTTGAGGCGTTGTGCCAGACCTTGGAAACC   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCCGGCGCTACAGCCAGCCTCACGACGACGGGCGCCGCGTTGAGGCGTTGTGCCAGACCTTGGAAACC   180

Qy   181 TCAGGTATATACCTTTTCCAGACGGCGGGGATCTCCCCTCCCCCATCCATAGTGCCTTGGGAC   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCAGGTATATACCTTTTCCAGACGGCGGGGATCTCCCCTCCCCCATCCATAGTGCCTTGGAC   240

Qy   241 CAAATCCAGGGCCGTTCTTTCCAGGAAACAATGAAGGGAGAACAGCAGACATCTGAATGAAGA   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAAATCCAGGGCCGTTCTTTCCAGGAAACAATGAAGGGAGAACAGCAGACATCTGAATGAAGA   300

Qy   301 AGAGGGTGCCAGCGGGGTATGAGGAGTGCATTATCGTTATAGTGGGAACTTCAGTGACCAGTC   360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGAGGGTGCCAGCGGGGTATGAGGAGTGCATTATCGTTATAGTGGGAACTTCAGTGACCAGTC   360

Qy   361 CTCAGACACGAAAGGATGCTCCCTCACCCCCAGTCTTGGAGGGCAAATCTGCACAGAGCCAGT   420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTCAGACACGAAAGGATGCTCCCTCACCCCCAGTCTTGGAGGGCAAATCTGCACAGAGCCAGT   420

Qy   421 CTGCAACACCAGAGACCAGAGGGCCGCCAGGTCAAGCTCCCGGGTGTGTCTAAGAGGGGAGGTCTC   480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTGCAACACCAGAGACCAGAGGGCCGCCAGGTCAAGCTCCCGGGTGTGTCTAAGAGGGGAGGTCTC   480

Qy   481 CAGCCTTCTGAATTACACGCACCGACACATGACACAGGAGATGGAGACAGAGAGATGATGAAGTAGA   540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAGCCTTCTGAATTACACGCACCGACACATGACACAGGAGATGGAGACAGAGAGATGATGAAGTAGA   540

Qy   541 TGATGGGAATGGCTCTGATAATTCTAATGCCAAAGCTCACCCGTGAGACCAAAGGACACCAG   600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TGATGGGAATGGCTCTGATAATTCTAATGCCAAAGCTCACCCGTGAGACCAAAGGACACCAG   600

Qy   601 GACGGCGCTCTGAAAAGCCCGGCGTGTCCGAAACCCGACATAGCAATGGGACCTCCAGCTTGGA   660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GACGGCGCTCTGAAAAGCCCGGCGTGTCCGAAACCCGACATAGCAATGGGACCTCCAGCTTGGA   660

Qy   661 GAGGCAAAGAGCCTCCCCCAGAATCACCCGAGGTCGGCGGGCGCCGCCACCATGTCCAGGA   720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GAGGCAAAGAGCCTCCCCCAGAATCACCCGAGGTCGGCGGGCGCCGCCACCATGTCCAGGA   720

Qy   721 GTACCCTGTGGGAGTTTCCGGGCTACCAGGTCTCGGAGAGACGTCGAGCATCGTCTTCAGCAAG   780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTACCCTGTGGGAGTTTCCGGGCTACCAGGTCTCGGAGAGACGTCGAGCATCGTCTTCAGCAAG   780

Qy   781 CACGCCAATGGTCATCCCCTGCCAGCCGTCAGCACTTCATGGGAGAGAAGTGCACACCTAAGAGCCGT   840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CACGCCAATGGTCATCCCCTGCCAGCCGTCAGCACTTCATGGGAGAGAAGTGCACACCTAAGAGCCGT   840

Qy   841 CAGTACCCCATCAGTTGACTTGAGCCAGGATGGCAGAGATCAGAGAGGGTATGGATACCACACA   900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CAGTACCCCATCAGTTGACTTGAGCCAGGATGGCAGAGATCAGAGAGGGTATGGATACCACACA   900

Qy   901 GGTGGAGTGCAGAGAGGAGCAGAGATGGAGACAGCAGCACAGAGTATCAGGATGATGAAAGAGTTTGG   960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GGTGGAGTGCAGAGAGGAGCAGAGATGGAGACAGCAGCACAGAGTATCAGGATGATGAAAGAGTTTGG   960

Qy   961 AATAGGGTGACCTCGTGTGGGGAAAGATCAAGGGGCTTCTCCTGGGTGGCCCTGCCAATGGTGGT   1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AATAGGGTGACCTCGTGTGGGGAAAGATCAAGGGGCTTCTCCTGGGTGGCCCTGCCAATGGTGGT   1020

Qy  1021 GTCCTGGAAAAGCCACCTCCAAGCGACAGGACCACATGCCCGGAAATGCCCTGGGTACAGTGGTT   1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 GTCCTGGAAAAGCCACCTCCAAGCGACAGGACCACATGCCCGGAAATGCCCTGGGTACAGTGGTT   1080

Qy  1081 TGGTGATGGCAAGTTTTCTGGATCTCTCGTCTTGACAAAACTGGTGGCCTGGGCGTTTCAG   1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 TGGTGATGGCAAGTTTTCTGGATCTCTCGTCTTGACAAAACTGGTGGCCTGGGCGTTTCAG   1140

Qy  1141 CCAGCACTTTAATCTGGCTACCTTCAATAAGCCTGGTTTTCTTATAGGAGAGGCCATGTTACCA   1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CCAGCACTTTAATCTGGCTACCTTCAATAAGCCTGGTTTTCTTATAGGAGAGGCCATGTTACCA   1200

Qy  1201 CACTCTGGAGAAAAGCCAGGGTTCGAGCTGGCAAGAACCTTCTCCAGCAGTCCTGGAGAGTC   1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CACTCTGGAGAAAAGCCAGGGTTCGAGCTGGCAAGAACCTTCTCCAGCAGTCCTGGAGAGTC   1260
```

```
Qy  1261 ACTGGAGGACCAGCTGAAGCCCATGCTGGGAGTGGGGGGCTTCAAGCCTACTGG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 ACTGGAGGACCAGCTGAAGCCCATGCTGGGAGTGGGGGGCTTCAAGCCTACTGG 1320

Qy  1321 GATCGAGGGGCCTCAAACCCAACAAGAAGCAACCAGTGGGCCATCAA 1380
         ||||||||||||||||||||||||||||||||||||||||
Db  1321 GATCGAGGGGCCTCAAACCCAACAAGAAGCAACCAGTGGGCCATCAA 1380

Qy  1381 TTCAGACAGTAGGGAACTTAGAACCCAGGAGACGGCACAAC 1440
         ||||||||||||||||||||||||||||||||||||
Db  1381 TTCAGACAGTAGGGAACTTAGAACCCAGGAGACGGCACAAC 1440

Qy  1441 CAATGACTCTGCTTCTGAGTCCCCCACCCAAAATAGCTATGG 1500
         ||||||||||||||||||||||||||||||||||
Db  1441 CAATGACTCTGCTTCTGAGTCCCCCACCCAAAATAGCTATGG 1500

Qy  1501 CGGGAAGGGACCGAGGGGGAGGAGGATGAGAGGAGGTCTGAAGTCACCAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 CGGGAAGGGACCGAGGGGGAGGAGGATGAGAGGAGGTCTGAAGTCACCAA 1560

Qy  1561 CAACAAGGGCAAATCTGGAAGACCGGCTGTTTTGTCCTGTGAAAGAGAACCCTGTTGTCCTT 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 CAACAAGGGCAAATCTGGAAGACCGGCTGTTTTGTCCTGTGAAAGAGAACCCTGTTGTCCTT 1620

Qy  1621 CCACCCCCTCTTTGAGGGTGGGGCTTCTGTCAGAGTTGCCGGGGATCGCTTCCTAGAGCTCTT 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 CCACCCCCTCTTTGAGGGTGGGGCTTCTGTCAGAGTTGCCGGGGATCGCTTCCTAGAGCTCTT 1680

Qy  1681 CTACATGTATGATGAGGACGGGCTATCAGTCCTACTGCACCGTGTGCTGTGTGAGGGCCGTGA 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 CTACATGTATGATGAGGACGGGCTATCAGTCCTACTGCACCGTGTGCTGTGTGAGGGCCGTGA 1740

Qy  1741 ACTGCTGGTGCAGTAACACAAGCTGCTGCAGATGCTTCTGTGTGTGGAGTGTCTGGAGGT 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 ACTGCTGGTGCAGTAACACAAGCTGCTGCAGATGCTTCTGTGTGTGGAGTGTCTGGAGGT 1800

Qy  1801 GCTGTGGGGGCCGCCACAGTGTGAGGATGCCCAAGCCTGACCTGGAAAGAATTTGACTCTAT 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 GCTGTGGGGGCCGCCACAGTGTGAGGATGCCCAAGCCTGACCTGGAAAGAATTTGACTCTAT 1860

Qy  1861 GTGCCTCCCTCAGCGCCTGCCATGGGGGTCCTCCGACGCAGGAAAGATTGGAACATGCGCCT 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 GTGCCTCCCTCAGCGCCTGCCATGGGGGTCCTCCGACGCAGGAAAGATTGGAACATGCGCCT 1920

Qy  1921 GCAAGACTTCTTCACTCATGATCCTGACCTGGAAGAAATTTGAGCCACCCAAGTTGTACCC 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 GCAAGACTTCTTCACTCATGATCCTGACCTGGAAGAAATTTGAGCCACCCAAGTTGTACCC 1980

Qy  1981 AGCAAATTCCTGCAGCCAAAAAGGAGGGCCCATTAGAGTCCTGGTATTAAAGTGGAAAAGTACA 2040
         ||||||||||||||||||||||||||||||||||||||||||||
Db  1981 AGCAAATTCCTGCAGCCAAAAAGGAGGGCCCATTAGAGTCCTGGTATTAAAGTGGAAAAGTACA 2040

Qy  2041 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTGGAAAAGTACATTGCCTCCGA 2100
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTGGAAAAGTACATTGCCTCCGA 2100

Qy  2101 AGTCTGTGCAGAGTCCATCGCGTGTGGGGAACTGTTAAGCATGAAGGCCAGATCAAAATATGT 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 AGTCTGTGCAGAGTCCATCGCGTGTGGGGAACTGTTAAGCATGAAGGCCAGATCAAAATATGT 2160

Qy  2161 CAATGACGTCCGGAAAATCACCAAGAAAAAATATTGAAGAGTGGGGCCGTTCGACTTGGT 2220
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CAATGACGTCCGGAAAATCACCAAGAAAAAATATTGAAGAGTGGGGCCGTTCGACTTGGT 2220

Qy  2221 GATTGGTGGAAGCCCATGCAATGATCTCTCTAAGTCAATCCTGCCCGCAAAGGTTTATA 2280
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2221 GATTGGTGGAAGCCCATGCAATGATCTCTCTAAGTCAATCCTGCCCGCAAAGGTTTATA 2280

Qy  2281 TGAGGGCACAGGAAGGCTCTTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCCGCCCAA 2340
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 TGAGGGCACAGGAAGGCTCTTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCCGCCCAA 2340

Qy  2341 GGAGGGCGCACAACCGTCCAATTCTTCTGGAATGTTGTGCGGCCATGAGAAAAGTGAA 2400
         ||||||||||||||||||||||||||||||||||||||||
Db  2341 GGAGGGCGCACAACCGTCCAATTCTTCTGGAATGTTGTGCGGCCATGAGAAAAGTGAA 2400
```

```
RESULT 2
  1969
S
```

```
Qy  3421  TGGCACCTTCCCTTGCTTTGGTTGGTACAAAGGGCTGAAGTCCTGTTGGTTCTTGTAGCATTTCC  3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  CAGGATGAATGATGTCAGCAGGGGATGACATCACCACCTTTAGGCTTTCCCTGGCCAGGG  3540

Qy  3481  CAGGATGAATGATGTCAGCAGGGGATGACATCACCACCTTTAGGCTTTCCCTGGCCAGGG  3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  CCCATGTGGCTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAAGGGTGGG  3600

Qy  3541  CCCATGTGGCTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAAGGGTGGG  3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  TGGAGTGGCCCCTCTTCCAGGTGTGAGGGATACGGAAGGAGGAAGCTTAGGGAAAATCCAATTC  3660

Qy  3601  TGGAGTGGCCCCTCTTCCAGGTGTGAGGGATACGGAAGGAGGAAGCTTAGGGAAAATCCAATTC  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  CCCACTCCCTCTGTTGCCAAATCAGGGCCCCAGTCCCCAAACAGCTCAGGTCCCCACAAACCCC  3720

Qy  3661  CCCACTCCCTCTGTTGCCAAATCAGGGCCCCAGTCCCCAAACAGCTCAGGTCCCCACAAACCCC  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  CTAGTTCCTCATGAGAAGCTAGGACCACATCGTTCCCCTTATCTGAGCAGTGTT  3780

Qy  3721  CTAGTTCCTCATGAGAAGCTAGGACCACATCGTTCCCCTTATCTGAGCAGTGTT  3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781  TGGGGAACTACAGTGAAAAACCTTCTGGAGATGTTAAAAGCTTTTTACCCCACGATAGATT  3840

Qy  3781  TGGGGAACTACAGTGAAAAACCTTCTGGAGATGTTAAAAGCTTTTTACCCCACGATAGATT  3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841  GTGTTTTTAAGGGGTGCCTTTTTTTAGGGCATCACTGCAATCACACTGCATTTCAG  3900

Qy  3841  GTGTTTTTAAGGGGTGCCTTTTTTTAGGGCATCACTGCAATCACACTGCATTTCAG  3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901  AAATGCCATCGTAAATGGTTTTTAAACACCTTTACCTAATTACAGGTGCTATTTTTATAGA  3960

Qy  3901  AAATGCCATCGTAAATGGTTTTTAAACACCTTTACCTAATTACAGGTGCTATTTTTATAGA  3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961  AGCAGACAACACTTCTTTTTATGACTCTCACACTTCTATTTTCATGTTACCCATTTTTTTT  4020

Qy  3961  AGCAGACAACACTTCTTTTTATGACTCTCACACTTCTATTTTCATGTTACCCATTTTTTTT  4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  GTAACTCGCAAGGTGTGGGGCTTTTTGTAAACTTCACCAGGTGTGTGGGAGAGACTGCCTTGTTT  4080

Qy  4021  GTAACTCGCAAGGTGTGGGGCTTTTTGTAAACTTCACCAGGTGTGTGGGAGAGACTGCCTTGTTT  4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081  CAACAGTTTGTCTCCCACTGGTTTCTAATTTTTAGGTGCAAAGATCACACAGATGCCCAGAGT  4140

Qy  4081  CAACAGTTTGTCTCCCACTGGTTTCTAATTTTTAGGTGCAAAGATCACACAGATGCCCAGAGT  4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4141  TTACCTTTCTGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAA  4195

Qy  4141  TTACCTTTCTGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAA  4195
```

```
DEFINITION  Mus musculus DNA cytosine-specific methyltransferase isoform 1
            (Dnmt3b) mRNA, complete cds.
ACCESSION   AF151969            4338 bp    mRNA    linear    ROD 08-JUN-2000
VERSION     AF151969.1  GI:8347117
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4338)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE     Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4338)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences; Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China
FEATURES             Location/Qualifiers
     source          1..4338
                     /organism="Mus musculus"
                     /strain="KM"
                     /db_xref="taxon:10090"
                     /dev_stage="8-9 day old embryo"
     gene            1..4338
                     /gene="Dnmt3b"
     CDS             412..2991
                     /gene="Dnmt3b"
                     /function="methylates cytosine in DNA"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA cytosine-specific methyltransferase isoform
                     1"
                     /protein_id="AAF74515.1"
                     /db_xref="GI:8347118"
                     /translation="MKGDSRHLNEEBGASCYEECIIVNGHNFSDQSSDTKDAPSPPVLE
                     AICTEPVCTPETRGRSRSSSRLSKREVSSLLNYTQDMTGDGRDRDEVDDNGSSDIILMPK
                     LTRETKDTRRTRSESPAVRTRHSNGTSSLEQRASPRITRCGROGRHHVDEYPVEPATR
                     SRRRASSASSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDQEGMDTTQVDAESIY
                     GDSTEYQDDKEFGIGDLVWGKIKGFSWWPANVVSWKATSKRQAMPGHRVWQWFGDGKF
                     SEISADKLVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTPSSSSPGESLED
                     QLKPMLEWAHGGFKPTGIEGLKPNNKQPVVNKSKVRRSDSRNLEPRRENKSRRRTTN
                     DSAASESPPPRRLKTNSYGKDRGEDEDEESRERMASEVTNKRNGNLEDRCLSCGKKNPVS
                     FHPLEGGLCQSCRDRFLELFYMIDEDGYQSYCTVCCEGRELLLCSNTSCCRCPCVEC
                     LEVLVGAGTAEDAKLQEPWSCYMCLPQRCHGVLRRRKDWNMRLQDFFTTPDLEEFEP
                     PKLYPAIPAARRRPINVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKH
                     EGQIKIVNDVRKITKKNIEEWGFPDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEEY
                     HLLNYTRPKEGDNRPFFWMFENVVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARY
                     FWGNLPGMNRPVMASKNDKLELQDCLEFSRTAKLKKVQTITTKSNSIROGKNQLFPVV
                     MNGKDDVLWCTELERIFGFPAHYTDVSNMGRGAROKLLGRSWSVPVIRHLFAPLKDYF
                     ACE"
BASE COUNT      1073 a   1124 c   1173 g   968 t
ORIGIN
```

```
Query Match        99.7%;   Score 4182.8;   DB 10;   Length 4338;
Best Local Similarity  100.0%;   Pred. No. 0;   Mismatches 2;  Indels 0;  Gaps 0;
Matches 4184;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

Qy    10  GCGCGCGGGGTTAAGCGCGGCCCAAGTAAACGTACGGCACGCATCGGCCGCCCCGAGAGATTCGGCGA  69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   153  GCGCGCGGGGTTAAGCGCGCCCAAGTAAAACGTACGGCACGCATCGGCCGCCCGAGAGATTCGGCGA  212

Qy    70  ACCCGACACTCCGCGCGCGCGCCGCGGCCGCGCCCCGCGCGGGCCCGGCGCCGCCCCGCGCTA  129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70  ACCCGACACTCCGCGCGCGCGCCGCGGCCGCGCCCCGCGCGGGCCCGGCGCCGCCCCGCGCTA  129

Qy   130  CAGCCAGCCTCACGACACAGGCCCGCTGAGGCTTGTGCCAGACCTTTGGAAACCTTCAGGTATA  189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   273  CAGCCAGCCTCACGACACAGGCCCGCTGAGGCTTGTGCCAGACCTTTGGAAACCTTCAGGTATA  332

Qy   190  TACCTTTCCAGACGCGGGGAATCTCCCCTCCCCATCCATAGTGCCCTTGGGACCAAAATCCAG  249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   333  TACCTTTCCAGACGCGGGGAATCTCCCCTCCCCATCCATAGTGCCCTTGGGACCAAAATCCAG  392

Qy   250  GGCCTTCTTTTCCAGGAAACACAAATGAAAGGGAGACAGCAGCACATCGAATGAAAGAAGAAGAGGGTGC  309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   393  GGGCCTTCTTTTCCAGGAAAACAAATGAAAGGGAGACAGCAGCACATCGAATGAAAGAAGAAGAGGGTGC  452

Qy   310  CACGCGGGTRATGAGGAGTGCCATTATACGTTAAATGGGAACTTCAGTGACCAGTCCTCAGACAC  369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   453  CAGCGGGTRATGAGGAGTGCATTATCCGTTAAATGGGAACTTCAGTGACCAGTCCTCAGACAC  512

Qy   370  GAAGGATGCTCCCTCACCCCCAGTCTTGGAGGCAAATCTGCACAGACAGCCAGTCTGCACACC  429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   513  GAAGGATGCTCCCTCACCCCCAGTCTTGGAGGCAAATCTGCACAGACAGCCAGTCTGCACACC  572

Qy   430  AGAGACCAGAGGCCCGCAGGTCAAGCTCCGGCTGTCTAAGAGGAGGAGGTCTCCAGCGCTTCT  489
```

```
Db    573 AGAGACCAGAGGCCGCAGGTCAAGCTCCCGGCTGTCTAAGAGGAGGCTCTCCAGCTTCT 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    .

Db    490 GAATTACACGCAGGACATGACAGGAGAGATGGAGACAGAGATGAAGTAGATGATGGGAA 549
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Qy    633 GAATTACACCCAGGACATGACAGGAGAGATGGAGACAGAGATGAAGTAGATGATGGGAA 692

Qy    550 TGGCTCTGATATTCTAATGCCAAAAGCTCACCCTGAGACCACCCAGGACGCGCGCCTC 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 TGGCTCTGATATTCTAATGCCAAAAGCTCACCCTGAGACCACCCAGGACGCGCGCCTC 752

Qy    610 TGAAAGCCCGGCTGTCCGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAAG 669
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 TGAAAGCCCGGCTGTCCGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAAG 812

Db    670 AGCCTCCCCCAGAATCACCCAGAGGTCGGCCACCATGTCCAGGAGTACCCCTGT 729
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    813 AGCCTCCCCCAGAAATCACCCAGAGGTCGGCCACCATGTCCAGGAGTACCCTGT 872

Qy    730 GGGAGTTCCGGCTACCAGGTCTCGGAGACGTCGAGCACTCCAGGAGCACGCGCCATG 789
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 GGGAGTTCCGGCTACCAGGTCTCGGAGACGTCGAGCACTCCAGGAGCACGCGCCATG 932

Db    790 GTCATCCCCTGCCCAGGGTCGCACTTCATGGAAGAAGTGACACCTAAGACGCGTCAGTACCCC 849
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    933 GTCATCCCCTGCCCAGGGTCGCACTTCATGGAAGAAGTGACACCTAAGACGCGTCAGTACCCC 992

Qy    850 ATCAGTTGACTTGAGCCAGGATCAGAGAATCAGAGAGAGGGTATGGAGATCAGAGGAG 909
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GGAGTTTGACTTGAGCCAGGATCTCGGAGACGTCCAGGAGCATCGTCTTCAGCAAGCAC 1052

Db    910 AGAGAGCAGAGAATGGAGAACAGCAGCAGTATCAGGAGTGATGATAAAAGAGTTTGGAATAGGTGA 969
          |||||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Qy    993 ATCAGTTGACTTGAGCCAGGATCAGAGAATCAGAGAGAGGGTATGGAGATCAGAGGAGGGATGC 1052

Qy    970 CCTCGTGTGGGGGAAAGATCAAGGGCCTTCTCCTGCTGTGGGGCCTGCGCTGGTGTGTCCTGGAA 1029
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 AGAGAGCATATATGGAGACAGCAGCAGTATCAGGAGTGATGATAAAAGAGTTTGGAATAGGTGA 1112

Db   1030 AGCCACCTCAAGCGACAGGCCATGCCCCGGAATGCCGCTGGGGTACAGTCGGTTTGGTGATGG 1089
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1113 CCTCGTGTGGGGGAAAGATCAAGGGCCTTCTCCTGCTGTGGGGCCTGCGCTGGTGTGTCCTGGAA 1172

Qy   1150 TAATCTGGCTACCTTCAATAAGCTGGTTCTTTATAGGAGGCCCATGTACACCACACTCTGA 1209
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 AGCCACCTCCAAGCGACAGGCCCATGCCCCGGAATGCGCTGGGTACACTGGTTTGGTGATGG 1232

Db   1210 CAAGTTTCTGAGATCTCGCTGGTCAGGTGGCTCTCTGGGCGCTGTTCAGCCAGCACTT 1269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1233 CAAGTTTTCTGAGATCTCTGCTCGACAAACTGGTGGCTGCTCTGGGGGTGTTCAGCCAGCACTT 1292

Qy   1210 GAAAGCCAGGGTTCGAGCTGGCCAAGAACCTTCTCCAGCAGTCCACTGGAGGA 1269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 TAATCTGGCTACCTTCAATAAGCTGGTTCTTTATAGGAGGCCCATGTACACCACACTCTGGA 1352

Db   1353 GAAAGCCAGGGTTCGAGCTGGCCAAGAACCTTCTCCAGCAGTCCACTGGAGGA 1412
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1350 CAACCGTCCAATCTTCTTCGGATGGTTCGGAGAATGTTCGGGCCATGAAAAGTGAATGACAAGAA 1409
```

RESULT 3

```
AF068627
LOCUS       AF068627    4135 bp    mRNA    linear    ROD 06-DEC-1999
DEFINITION  Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF068627
VERSION     AF068627.2  GI:6449471
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4135)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE     Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
  JOURNAL   Nat. Genet. 19 (3), 219-220 (1998)
  MEDLINE   98324766
  PUBMED    9662389
REFERENCE   2  (bases 1 to 4135)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REFERENCE   3  (bases 1 to 4135)
  AUTHORS   Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK      Sequence update by submitter
COMMENT     On Nov 18, 1999 this sequence version replaced gi:3327981.
FEATURES             Location/Qualifiers
     source          1..4135
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="84.0 cM"
     gene            1..4135
                     /gene="Dnmt3b"
     misc_feature    1..4135
```

[This page is a two-column nucleotide sequence alignment between two sequences labeled "Qy" (query) and "Db" (database), with position numbers flanking each aligned row and vertical bars (|) indicating matching bases. The alignment spans positions approximately 1561 through 3840.]

```
Qy  3841 GTGTTTTTAAGGGGTGCTTTTTTTTTTAGGGGCATCACTGGAGATAACAAAGCTGCCATTTCAG 3900
         ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  3781 GTGTTTTTAAGGGGTGCTGCTTTTTTTTTTAGGGGCATCACTGGACATAAGAAAGCTGCCATTTCAG 3840

Qy  3901 AAATGCCATCGTAAATGGTTTTTTAAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGA 3960
         ||||||||||||||||||||||||  ||||||||||||||||||||||||||||||| ||||
Db  3841 AAATGCCATCGTAAATGGTTTTTTAAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGA 3900

Qy  3961 AGCAGACAACACTTCTTTTTATGACTTCTCAGACTTCTTATTTCATGTTACCATTTTTTTTT 4020
         ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  3901 AGCAGACAACACTTCTTTTTATGACTTCTCAGACTTCTTATTTCATGTTACCATTTTTTTTT 3960

Qy  4021 GTAACTCGCAAGGTGTGGGGCTTTTTGTAACTTCACAGGTGTGTGGGGAGACTGCCTTGTTT 4080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961 GTAACTCGCAAGGTGTGGGGCTTTTTGTAACTTCACAGGTGTGTGGGGAGACTGCCTTGTTT 4020

Qy  4081 CAACAGTTTGTTCTCCACTGGTTTCTTTCTAATTTTTTAGGTGCAAAGATGACAGAATGCCCAGAGT 4140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021 CAACAGTTTGTTCTCCACTGGTTTCTTTCTAATTTTTTAGGTGCAAAGATGACAGAATGCCCAGAGT 4080

Qy  4141 TTACCTTTTTCTGGTTGATTAAAGTTGTATTTCTTCTTAAAAAAAAAAAAAAAAA 4195
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081 TTACCTTTTCTGGTTGATTAAAGTTGTATTTCTTCTAAAAAAAAAAAAAAAAA 4135
```

RESULT 4

LOCUS   AF151970   4278 bp   mRNA   linear   ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 2
            (Dnmt3b) mRNA, complete cds.
ACCESSION   AF151970
VERSION   AF151970.1 GI:8347119
KEYWORDS   .
SOURCE   Mus musculus.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4278)
  AUTHORS   Yin,B., Chen.Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE   Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 4278)
  AUTHORS   Yin,B., Chen.Y.T., Zhu,M., Luo.Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China
FEATURES   Location/Qualifiers
     source   1..4278
                     /organism="Mus musculus"
                     /strain="KM"
                     /db_xref="taxon:10090"
                     /dev_stage="8-9 day old embryo"
     gene   1..4278
                     /gene="Dnmt3b"
     CDS   412..2931
                     /gene="Dnmt3b"
                     /function="methylates cytosine in DNA"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA cytosine-specific methyltransferase isoform
                     2"
                     /protein_id="AAF74516.1"
                     /db_xref="GI:8347120"
                     /translation="MKGDSRHLNEEEAGSAGYEECIIVNGNFSDSSDTKDAPSPVLE
                     AICEPVCTPETRGRRSSSRLSKREVSSLLNYTQDMTGDGDRDDEVDDGNGSDILMPK
                     SRRRASSASTPWSSPASVDPMEEVYPKSVSYPSVDLSODGDGEGMDTTQVDAESIY
                     GDSTEYQDDKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF"

BASE COUNT   1055 a   1113 c   1154 g   956 t
ORIGIN

Query Match   96.6%; Score 4052.8; DB 10; Length 4278;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4124; Conservative 0; Mismatches 2; Indels 60; Gaps 1;

```
Qy   10 GCGGCCGGGGGTTAAGCCGCCCCAAGTAAAACGTAGCCGCCAGCGAGCCGGGGGACGATTCCGGA  69
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCGGCCGGGGGTTAAGCGGGCCCAAGTAAAACGTAGCCGCCGATCGGGGGCCGGGGGAGATTCGGGA  69

Qy   70 ACCCGACAACTCCGCGCGCGCGCGCCCAGGAACCCGGGGGCGCGGGGGCGCGGGCCGCTA 129
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 ACCCGACACTCCCGCGCGCGCGCGCCCAGGGACCCGGGGGCGCGGGGGCGCGGGCCGCTA 129

Qy  130 CAGCCAGCCTCACGACACAGGCCCGCCGCTGAGCGCTTGTTCCCAGACCTTGGAAACTCAGGTATA 189
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  213 CAGCCAGCCTCACGACACAGGCCCGCCGCTGAGCGCTTGTTCCCAGACCTTGGAAACTCAGGTGC 272

Qy  190 TACCTTTCCAGACGCGGGGATGTCTCCCCTCCCCCATCCATAGTGCCTTGGGGACCAAATCCAG 249
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  333 TACCTTTCCAGACGCGGGGATCTCCCCTCCCCCATCCATAGTGCCTTGGGGACCAAATCAG 392

Qy  250 GGCCTTCTTTTCAGGAAAACAATGAAGGGAGACAGCAGCAGACAACTGAATGAAGAAGAGGGTGC 309
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  393 GGCCTTCTTTTCAGGAAAACAATGAAGGGAGACAGCAGCAGACAATCGAATGAAGAAGAGGGTGC 452

Qy  310 CAGCGGGTATGAGGAGTGCATTATCGTTAATGGGAACTTCAGTGACCAGTCCTCAGACAC 369
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  453 CAGCGGGTATGAGGAGTGCATTATCGTTAATGGGAACTTCAGTGACCAGTCCTCAGACAC 512

Qy  370 GAAGGATGCTCCCTCACCCCCCAGTGTTCTTGGAGCAATCTGCAGAGCCAGTCTGCACC 429
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  513 GAAGGATGCTCCCTCACCCCCAGTCTTGGAGCAATCTGCACAGAGCCAGTGCACACC 572

Qy  430 AGAGACCAGAGGCCGCAGGTCAAGCTCCCGGCTGTTGTCTAAGAGGAGGAGGTCTCCAGGCTTCT 489
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  573 AGAGACCAGAGGCCGCAGGTCAAGCTCCCGGCTGTCTAAGAGGAGGAGGTCTCCAGGCTTCT 632

Qy  490 GAATTACACGCAGGACAATGACAGGAGGATGGAGACAGAGATGAGGAGATAGAGATGGGAA 549
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  633 GAATTACACGCAGGACAATGACACAGGAGGATGGAGACAGAGATGAGGAGATAGAGATGGGAA 692

Qy  550 TGGCTCTGATAATTCTAATGCCAAAGCTCACCCGTGAGACCAAGGACACCAGGACGGCGCTC 609
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  633 TGGCTCTGATAATTCTAATGCCAAAGCTCACCCGTGAGACCAAGGACACCAGGACGGCGCTC 609

Qy  610 TGAAACCCGGCGTGTCCGAACCCGACACATAGCAAATGGGACCTCCCAGCTTGGAGGGCAAAG 669
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  693 TGGCTCTGATATTCTAATGCCAAAGCTCACCCGTGAGACCAAGGACACCAGGACGGCGCTC 752

Qy  670 AGCCTCCCCAGAATCACCCGAGGTCGGCAGGGCGCGCCACCAACGTGCAGGAGGAGTACCCTGT 729
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  753 TGAAAACCCGGCGTGTCCGAACCCGACACATAGCAAATGGGACCTCCCAGCTTGGAGGGCAAAG 812

Qy  730 GGAGTTTCCGGCTACCAAGGTTCTTCGGGAGCACGTGCGGACCATCTGTCTTCAGCAACGGCCATG 789
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  813 AGCCTCCCCAGAATCACCCGAGGTCGGCAGGGCGCGCCACCAACGTGCAGGAGGAGTACCCTGT 872

Qy  790 GTCATCCCCTGCCAGCCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCC 849
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  873 GGAGTTTCCGGCTACCAAGGTTCTTCGGGAGCACGTGCGGACCATCTGTCTTCAGCAACGGCCATG 932

Qy  790 GTCATCCCCTGCCAGCCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCC 849
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  933 GTCATCCCCTGCCAGCCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCC 992
```

Query Match protein (translated):

```
SEISADKLYLALGLFSQHFNLATFENKLVYSYRKAMYHTLEKARVRAGKTFSSSPGESLED
QLKPMLEWAHGGFKPTGIEGLKPNKKQPENKSRRRTTNDSAASESPPFRLKTNSTGG
KDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVSHPLFEGGLCQSCRDFLEL
FYMYDEDGYOSYCTVCCEGRELLLCSNTSCCCRCFVECLEVLVGAGTAEDAKLQEPWS
CYMCLPQRCHGVLRRKRDWNMRLQDFFTDPDLEEFEPPRKLYPAIPAAKRPIRVLSL
FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYVNDVRKITKKNIEE
WGPPDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYHLLNYTRPKEGDNRPFFWMF
ENVVAMKVNDKDISRFLACNPVMIDAIKVSAAHRARYFWGKDPVLWMASKNDKL
ELQDCLEFSRTAKLKKVQTITTKSNISIRQGKNQLFPVVWNGKDDVLWCIELERIFGFP
AHYTDVSNMGRGAROKLLGRSWSVPVIRHLFAPLKDYFACE"
```

```
Qy    850  ATCAGTTGACTTGACCCAGGATGGAGAGATCAGGAGGAGGGTATGGATACCACACAGGTGGATGC  909
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993  ATCAGTTGACTTGACCCAGGATGGAGAGATCAGGAGGAGGGTATGGATACCACACAGGTGGATGC  1052

Qy    910  AGAGAGCAGAGATGGAGAGACCACGACAGAGTATCAGGACGTATCAGGAGGGTTTGGAATAGGTGA  969
           ||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053  AGAGAGCAATATATGGGGACACACAGAGTATCAGGACGATGAATAAAGAGTTTGGAATAGGTGA  1112

Qy    970  CCTCGGTGTGGGGGAAAGATCAAGGGCTTCTCTCCTGGTGGGCCTCCATGGTGGTGTCCTGGAA  1029
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113  CCTCGTGTGGGGGAAAGATCAAGGGCTTCTCCTGGTGGGCCTGCCATGGTGGTGTCCTGGAA  1172

Qy   1030  AGCCACCTCCAAGCGACACAGGGCCATGCCCCGGAATGCCCTGGGTACAGTGGTTTGGGATGG  1089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173  AGCCACCTCCAAGCGACACAGGGCCATGCCCCTGGGTACAGTGGTTTGGGATGG  1232

Qy   1090  CAAGTTTTCTGAGATCTCTGCTGACAAAACTGGTGGCTCTGGGGGCTGTTCAGCCAGCACTT  1149
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233  CAAGTTTTCTGAGATCTCCTGCTGACAAAACTGGTGCCTCTGTTCAGCCAGCACTT  1292

Qy   1150  TAATCTGGCTACCTTCAATAAGCTGGTTTTCTTATAGGAAGGCCATGTACCACACACTCTGGA  1209
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293  TAATCTGGCTACCTTCAATAAGCTGGTTTTCTTATAGGAAGGCCATGTACCACACTCTGGA  1352

Qy   1210  GAAAGCCAGGGGTTCGAGTGGCCAAAGAACCTTCTCCAGCAGTGCCTCGGAGACTCACTGGAGGA  1269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353  GAAAGCCAGGGTTCGAGCTGGGCAAAGAACCTTCTCCAGCAGTGCACTGGAGGA  1412

Qy   1270  CCAGCTGAAGCCCCATGCTGCTGGAGTGGGGCCCACCGGTGGCTTCAAGCCTACTGGGATCGAGGG  1329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413  CCAGCTGAAGCCCCATGCTGGGAGTGGGGCCCACCGGTGGCTTCAAGCCTACTGGGATCGAGGG  1472

Qy   1330  CCTCAAAACCAACAACAAGAGAAACCACTGGTTTAATAAGTGGGTGCGTCGTTCAGACAG  1389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473  CCTCAAAACCAACAACAAGAGAAGCAACCA----------------------------------  1497

Qy   1390  TAGGAACTTAGAACCCAGGAGAGAGCCGGAAGACCGACCAACCAATGACTC  1449
           ||||||||||||||||||||||||||||||||||||||
Db   1498  ----------------GAGAGCAAAAGTCGGAAGAACAGACCACAACCAATGACTC  1532

Qy   1450  TGCTGCTTCTGAGTCCCCCCCCACCCAAAGGGCCTCAAGACAAAAATAGCTATAGGGCGGGGAAGGA  1509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533  TGCTGCTTCTGAGTCCCCCACCCAAAGCGCCTCAAGACAAAATAGCTATAGGGCGGGGAAGGA  1592

Qy   1510  CCGAGGGGGAGGATGAGGAGGACCGAGAGGACCGGATGGGCTTCTGAAGTGTCACCAACAACAAGGG  1569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1593  CCGAGGGGAGGATGAGGAGAGCCGAGACGGATGGCTTCTCGAAGTCACCAACAACAAGGG  1652

Qy   1570  CAATCTGGAAGAGCCGCTGTGTTTGTCCCTGTGGAAAGAAGAAACCCTGTGTCCTTCCACCCCCT  1629
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653  CAATCTGGAAGAGACCGCTGTGTTTGTCCTGTGTGAAAGAAGAAACCCTGTGTCCTTCCACCCCCT  1712

Qy   1630  CTTTGAGGGTGGGGGCTCTGTCACAGAGTTGCCGGGGATGCCGCTTCCTAGAGCTCTTCTACATGTTA  1689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713  CTTTGAGGGTGGGGTCTCTGTCACAGAGTTGCCGGGGATGCCGCTTCCTAGAGCTCTTCTACATGTA  1772

Qy   1690  TGATGAGGACGGGCTATCAGTCCTACTGCACCCGTGTGCCTGTGTCGAAGTCGGTGAAACTGCCTGCT  1749
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773  TGATGAGGACGGGCTATCAGTCCTACTGCACCGTGTGCCTGTGTCGAGGGCCCGTGAAACTGCCTGCT  1832

Qy   1750  GTGCAGTAACACAAGCTGCTGCCAGATCGCTTCTGTGTGGAGTGCTGGAGGTGCTGGTGGGG  1809
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833  GTGCAGTAACACAAGCTGCTGCCAGATGCTTCTGTGTGGGGAGTGCTGGAGGTGCTGGTGGGG  1892

Qy   1810  CGCCAGGCACAGCTGAGGATGCCAAGCTGCAGGAACCCTGGAGCTGCTACTATATGTGCCTCCC  1869
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1893  CGCCAGGCACACGCTGAGGATGCCCAACCTGCCAGGAACCCTGAGAGCTGCTACTATATGTCCCTCC  1952

Qy   1870  TCAGCGCTGCCATGGGGTCTCCTCCGACGCACGGAAAGAGGAACATGCCGCCCTGCAAGGACTT  1929
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953  TCAGCGCTGCCATGGGGTCTCCTCCGACGCACGGAAAGAGGATTGAACATGCCGCCCTGCAAGGACTT  2012

Qy   1930  CTTCACTCACTGATCCTGACCTGGGAAGAATTTGAGCCACCCAAGTTGTACCCAGCGAATTCC  1989
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2013  CTTCACTCACTGATCCCTGACCTGGGAAGAATTTGAGCCACCCAAGTTGTACCCAGCGAATTCC  2072
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1990  TGCAGCCAAAAGGAGGCCCATTAGAGTCCTGTGTCTCTGTTTTGCAAGGAATTGCAACGGGGTA  2049

Db   2073  TGCAGCCAAAAGGAGGCCCATTAGAGTCCTGTGTCTCTGTTTTGCAAGGAATTGCAACGGGGTA  2132
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2050  CTTGGTGCTCAAGGAGGTTGGGTATTAAAGTGGAAAAAGTAACAATTGCCTCCGAAGTCTGTGC  2109

Db   2133  CTTGGGTGCTCAAGGAGGTTGGGTATTAAAGTGGAAAAAGTAACAATTGCCTCCGAAGTCTGTGC  2192
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2110  AGAGTCCAATCGCTGTGTGGGAACTGCTTAAGCAATGGAAGGCCAGATCAAATAATGTCAATGACGT  2169

Db   2193  AGAGTCCATCGCTGTGTGGGAACTGCTTAAGCAATGGAAGGCCAGATCAAATAATGTCAATGACGT  2252
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2170  CCGGAAAATCACCAAGAAAAATAATTGAAGAGACTGGGGGCGCCCGGTTCGACTTGCTTGGTGATTGGTCGG  2229

Db   2253  CCGGAAAATCACCAAGAAAAATAATTGGAAGAGAGTGGGGCGCCCGGTTCGACTTGGTGATTGGTGG  2312
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2230  AAGCCCCATGCAATGACTTCTCTCTAACCTCAATCCCTGCCCGCAAAAGTTTATTATGAGGGCAC  2289

Db   2313  AAGCCCATGCAATGACTCTCTCTAACGTCAATCCTGCCCGCAAAAGTTTATTATGAGGGCAC  2372
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2290  AGGAAGGCTCTGTTCTCGAGTTTTTACCACTTGCTGAATTATTACCCGCCCCAAAGGAGGGGCGA  2349

Db   2373  AGGAAGGCTCTTCTTCCGAGTTTTTACCACTTGCTGAATTAACCCGCCCCAAAGGAGGGGCGA  2432
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2350  CAACGTCCATTCTTCTTCTGGATGTTGTGGCCATGAAAAGTGTGTGCCGGCATGAAAGTGAATGACAAGAA  2409

Db   2433  CAACCGTCCATTCTTCTTCTGGATGTTGCAGGACTGCCGGCTGGAGTTCAGTGAGGACAGACAAGAA  2492
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2410  AGACAATCTCAAGATTCCTGGCATGTAACCCAGTGATGATCGAATGCCATCAAGGTGTCTCGC  2469

Db   2493  AGACATCTCAAGATTCCTGGCATGTAACCCAGTGATGATCGAATGCCATCAAGGTGTCTCGC  2552
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2470  TGGCTCACAGGGCGCCCGGGTACTTCTTGGGGTAACCTACCCGGAATGAACAGGGCCCGTGATGGC  2529

Db   2553  TGGCTCACAGGGCGCCCGGGTACTTCTTGGGGTAACCTACCCGGAATGAACAGGGCCCGTGATGGC  2612
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2530  TTCAAAGAAATGATAAAGCTCGAGCTGCAGGACTGCCTGGAGGTTCAGTGAGCACGAAAGTT  2589

Db   2613  TTCAAAGAAATGATAAAGCTCGACTGCAGGACTGCCTGGAGTTCAGTAGGACAGACAAAGTT  2672
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2590  AAAGAAAAGTGCAGACAAATAAACCACCAAGTCGAAACTCCATCAGACAGGGCAAAAACCAGCT  2649

Db   2673  AAAGAAAAGTGCAGACAAATAAACCACCAAGTCGAAACTCCATCAGACAGGGCAAAAACCAGCT  2732
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2650  TTTCCCTGTAGTCATGAATGGCCAAGGACGACGACGTTTTGTTGTGCACTGCAGTCGGAAAGGAT  2709

Db   2733  TTTCCCTGTAGTCATGAATGGCCAAGGACGACGACGTTTTGTTGTGCACTGCAGTCGGAAAGGAT  2792
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2710  CTTCGGCCTTCCCCTGCTCACTACACGGACGCTGTCCAAACATGGGCGCCCGGGGCCCCGTCAGAA  2769

Db   2793  CTTCGGCGTTCCCCTGCTCACTACACGGACGTGTCCAAACATGGGCGCCCGGGGCCCCGTCAGAA  2852
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2770  GCTTGCTTGGGCAGTTCCTGGAGTTGTACCGGGTCATCAGACACCTGTTTTTGCCCCCTTGAAGGA  2829

Db   2853  GCTTGCTTGGGCAGTTCCTGGAGTTGTACCGGGTCATCAGACACCTGTTTTTGCCCCCTTGAAGGA  2912
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2830  CTACTTTCCCTGTGAATAGTTCTACCCAGAGGACTCGGTGGGAGCTCTCGGTCAGAGCCAGTGCC  2889

Db   2913  CTACTTTGCCCTGTGAATAGTTCTACCCAGAGGACTCGGTGGGAGCTCTCGGTCAGAGCCAGTGCC  2972
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2890  CAGAGTCACCCCTCCCTGAAAGGCACCTCACCTGTCCCCTTTTTTAGCCTCACCTGTTGTTGGGG  2949

Db   2973  CAGAGTCACCCCTCCCTGAAAGGCACCTCACCTGTCCCCTTTTTTAGCCTCACCTGTTGTTGGGG  3032
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2950  CCTCACATCACTGTACCTCAGCTTTTCCTCCTCCTCAGTGGGAGCACAGCGCCTCCTGGGCCCTT  3009

Db   3033  CCTCACATCACTGTACCTCAGCTTTTCTCCTCCTCAGTGGGAGCACAGCGCCTCCTGGCCCTT  3092
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3010  GCAGGGGAGCCCCGGTGCTCCCCTGTGCACAGCTCAGACCGTGGCCTGCTTGTTTAGAGTAGC  3069
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   4150 TGGTTGATTAAAGTTGTTATTTTCTGTTAAAAAAAAAAAAAAAAAAAAAAA 4195
          ||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   4233 TGGTTGATTAAAGTTGTTATTTTCTGTTAAAAAAAAAAAAAAAAAAAAAAA 4278

Db   3093 GCAGGGGAGCCCCGGTGCTGCTCCCCGGTGCTGCACAGCTCGGACCTGGCTTGCCTTAGC 3152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3070 CCGGCAGTGGTGCTCTCATGTTCTCTTCTTACCCTGCGAAACTTTAAAACTTGAAGTAGTTAGTAGA 3129

Db   3153 CCGGGCATGGTGCTCTCATGTTCTCTTCTTACCCTGCGAAACTTTAAAACTTGAAGTAGTTAGTAGA 3212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3130 TGGGCTTTCTTTTACCCTCCTCGAGTTATCACTCAGAAGTGATGGCTAAGATACCAAAAAA 3189

Db   3213 TGGGCTTTCTTTTACCCTCCTGAGTTATCACTCAGAAGTGATGGCTAAGATACCAAAAAA 3272
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3250 TTCATGCTGCAAAATCACTTGAGATTTTGTTTTTTTAAGTAACCCGTGCTCCACATTTGCTG 3309

Db   3273 ACAAACAAAAACAGAAACAAAAAACCTCAACAGACTCTCTTAGTACTCAGG 3332
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3190 ACAAACAAAAACAGAAACAAAAAACCTCAACAGACTCTCTTAGTACTCAGG 3249

Db   3333 TTCATGCTGCAAAATCACTTGAGATTTTGTTTTTTTAAGTAACCCGTGCTCCACATTTGCTG 3392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3310 GAGGATGGCTATTGTGAATGTGGGCCTCAGGGGGCCAAAAAAAATTCC 3369

Db   3393 GAGGATGGCTATTGTGAATGTGGGCCTCAGGGGGCCAAAAAAAATTCC 3452
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3370 CCCTCTCCCCCCAGGAGTATTGAAGATGATTGTTTGTTTTGAGTCTTCCTGGCACCTT 3429

Db   3453 CCCTCTCCCCCCAGGAGTATTGAAGATGATTGTTTGTTTTGAGTCTTCCTGGCACCTT 3512
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3430 CCCCTTGCTTTTGGTACAAAGGGCTGAAGTCCTGTTGGTCTGTTCCTTAGGGTTTAGGGAGTCCCTGGCATTTCCCAGGATGAT 3489

Db   3513 CCCCTTGCTTTTGGTACAAAGGGCTGAAGTCCTGTTGGTCTGTTCCTTAGGGAGTCCCTGGCATTTCCCAGGATGAT 3572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3490 GATGTGCAGGGATGACATCACCACCTTTAGGCGTTTTCCCTGGCCAGGGGCCCATGTGG 3549

Db   3573 GATGTCAGGCAGGGATGACATCACCACCTTTAGGCGTTTTCCCTGGCCAGGGGCCCATGTGG 3632
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3550 CTAGTCCTCACGAAGACTGGGAGTAGGAATGTTTGGGAGCTCAGGAAGGGTGGGGTGGC 3609

Db   3633 CTAGTCCTCACGAAGACTGGGAGTAGAATGTTTGGGAGCTCAGGAAGGGTGGGGTGGC 3692
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3610 CCTCTTCCAGGTGTGAGGGGATACCGAAAGGAGGAAGCTTAGGAGAAATCCATTCCCCACTCCC 3669

Db   3693 CCTCTTCCAGGTGTGAGGGGATACCGAAAGGAGGAAGCTTAGGAGAAATCCATTCCCCACTCCC 3752
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3670 TCTTGCCAAATGAGAGGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCCCTAGTTCCT 3729

Db   3753 TCTTGCCAAATGAGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCCCTAGTTCCT 3812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3730 CATGAGAAGCTAGGACCAGAAGACACATCGTTCCCCTTATCTGAGCACAGTGTTTGGGGAACT 3789

Db   3813 CATGAGAAGCTAGGACCAGAAGCACACATCGTTCCCCTTATCTGAGCACAGTGTTTGGGGAACT 3872
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3790 ACAGTGAAAAACCTTCTGGAGATGTTAAAAGCTTTTACCCCACGATAGAATTGTTGTTTTTA 3849

Db   3873 ACAGTGAAAAACCTTCTGGAGATGTTAAAAGCTTTTACCCCACGATAGAATTGTTGTTTTTA 3932
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3850 AGGGGTGCTTTTTTTTAGGGGCATCACTCGAGATAAGAAAAGCTGCCATTTCAGAAAATGCCAT 3909

Db   3933 AGGGGTGCTTTTTTTTAGGGGCATCACTCGAGATAAGAAAAGCTGCCATTTCAGAAAATGCCAT 3992
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3910 CGTAATGGTTTTTTAAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGAAGCAGACAGACAA 3969

Db   3993 CGTAAGGTTTTTTAAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGAAGCAGACAA 4052
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3970 CACTCTCTTTTTATGACTCTCAGACTTCTCTATTTTCATGTTACCATTTTTTTTTGTACTCGC 4029

Db   4053 CACTCTCTTTTTATGACTCTCAGACTTCTCTATTTTCATGTTACCATTTTTTTTTGTACTCGC 4112
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4030 AAGGTGTGGGGCTTTGTGAACTTCACAGGTGTGTGGGGAGAGAGACTGCCCTTGTTGTTTCAACAGTTT 4089

Db   4113 AAGGTGTGGGGCTTTGTGAACTTCACAGGTGTGTGGGGAGAGAGACTGCCCTTGTTGTTTCAACAGTTT 4172
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4090 GTCTCCACTGGTTTCTAATTTTTTAGGTGCAAAAGATGACAGAGACCCAGAGTTTACCTTTC 4149
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4173 GTCTCCACTGGTTTCTAATTTTTTAGGTGCAAAAGATGACAGAGACCCAGAGTTTACCTTTC 4232
```

RESULT 5

Qy AF151973

Db AF151973

LOCUS   AF151973   4223 bp   mRNA   linear   ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 5
            (Dnmt3b) mRNA, complete cds.
ACCESSION   AF151973
VERSION   AF151973.1 GI:8347127
KEYWORDS   .
SOURCE   Mus musculus.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4223)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE   Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 4223)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China
FEATURES   Location/Qualifiers
     source   1..4223
                     /organism="Mus musculus"
                     /strain="KM"
                     /db_xref="taxon:10090"
                     /dev_stage="8-9 day old embryo"
     gene   1..4223
                     /gene="Dnmt3b"
     CDS   297..2876
                     /gene="Dnmt3b"
                     /function="methylates cytosine in DNA"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA cytosine-specific methyltransferase isoform
                     5"
                     /protein_id="AAF74519.1"
                     /db_xref="GI:8347128"
                     /translation="MKDSRHLNEEEGASGYEECIIVNGNFSDQSSDTKDRSPPVLE
                     AICTEPVCTPETRGRRSSSRLSKREVSSLLNYTQDMTGDCRDDEVDDGNGSDILMPK
                     LTRETKDTRTRSESPAVRTRHSNGTSSLERQRASPRITRGRQGRHHVQEYPVPPATR
                     SRRRASSSSASTPWSSPASVDMEEEVTRKSVSTRSVDLSQDGDDQESMDTTQVDAESIY
                     GDSTEYQDDKEPGIGDLVWGKIRGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                     SEISADKLVALGLFSQHFLSQHFNLATRNKLVSYRKAMYHTLEKARVRAGKTFSSSPGESLED
                     QLKPMLEWAHGGEFKPTGIEGLRKPNKKQPVVNKSKVRRSDSRNLEPRRENKSRRRTTN
                     DSAASESPPPKRLKTNSYGGKDRGEDEDSRSRERMASEVTNNKGNLEDRCLSCGRKNPVS
                     FHPLEGGLCQSCRDRLELFYMYDEDGYOSYCTVCCEGRELLLCSNTSCCRCFCVEC
                     LEVLVGAGTAEDAKLQEPWSCYMCLPQRCHGVLRRKDWNMRLQDFTTDPDLEEFEP
                     PKLYPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKH
                     EGQIKVNVDVKRITTKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFY
                     HLLNTYRPKEGGDNRPFFWMFENVVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARY
                     FWGNLPGMNRPVWASKNDKLELQDCLEFSRTAKLKKVQTITTKSNSIRQGKNQLFPVV
                     MNGKDDVLWCTELERIGFPAHTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYF
                     ACE"
BASE COUNT   1084 a   1147 c   941 g   941 t
ORIGIN

Query Match   94.0%; Score 3942.8; DB 10; Length 4223;
  Best Local Similarity   99.8%;
  Matches 3938; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

```
Qy   251 GCCTTCTTTCAGGAAACAAATGAAGGAGAACAGCAGACATCTGAATGAAGAAGAGGTGCC 310
```

```
Qy   2471 GCTCACAGGGCCGCCGGTACTTCTGGGGGTAACCTACCCGGAATGAAACGGCCCGTGATGGCT 2530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2499 GCTCACAGGGCCGCCGGTACTTCTGGGGGTAACCTACCCGGAATGAAACAGGCCCGTGATGGCT 2558

Qy   2531 TCAAAGAAATGATAAGCTGCAGGACTGCCTGGAGTTCAGGTAGGACAGCAGCAAAGTTA 2590
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2559 TCAAAGAAATGATAAGCTGCAGGACTGCCTGGAGTTCAGGTAGGACAGCAGCAAAGTTA 2618

Qy   2591 AAGAAAGTGCAGACAATAACCACCAAGTGGAACTCCATCAGACAGGGCAAAAACCAGCTT 2650
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2619 AAGAAAGTGCAGACAATAACCACCAAGTCGAACTCCATCAGACAGGGCAAAAACCAGCTT 2678

Qy   2651 TTTCCCTGTAGTCATGAAATGGCAAGGACAGCAGCACGCTTTTGGTGGCTCAGACACTA 2710
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2679 TTTCCCTGTAGTCATGAAATGGCAAGGACAGCACGTTTTTGGTGGCTCGAAAGGATC 2738

Qy   2711 TTCGGCTTCCCTGCTCACTACACGGACGGTGTCCAACATGGGGCGCGCGGCGCCCGGTCAGAAG 2770
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2739 TTCGGCTTCCCTGCTCACTACACGGACGGTGTCCAACATGGGGCGCGCGGCGCCCGGTCAGAAG 2798

Qy   2771 CTGCTGGGGCAGGTCCTGGAGTGGTACCGGTGTCATCAGACACACCTGTTTGCCCCCTTGAAGGAC 2830
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2799 CTGCTGGGGCAGGTCCTGGAGTGGTACCGGTGTCATCAGACACACCTGTTTGCCCCCTTGAAGGAC 2858

Qy   2831 TACTTTGCCTGTGAAATAGTTCTACCCGGAGGAGCGCTCTCGGGTTCAGAGGCCAGTGCCC 2890
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2859 TACTTTGCCTGTGAAATAGTTCTACCCGGAGGAGCGCTCTCGGGTTCAGAGGCCAGTGCCC 2918

Qy   2891 AGAGTCACCCCTCCCCTGAAAGGCACCTCACCTGTCCCCCTTTTTAGCTCACCTGTGTGTGGGGC 2950
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2919 AGAGTCACCCCTCCCCTGAAAGGCACCTCACCTGTCCCCCTTTTTAGCTCACCTGTGTGTGGGGC 2978

Qy   2951 CTCACATCACTGTACCTCAGCGTTTCTTCTCCTGCTCAGTGGGGAGCGAGCCCCCTTG 3010
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2979 CTCACATCACTGTACCTCAGCTTTCTTCTCCTGCTCAGTGGGGAGCGAGCCCCCTTG 3038

Qy   3011 CAGGGGAGCACCCGGTGCTCCCGTGTGCACAGCTCAGACCTGGCTGCTTAGAGTAGCC 3070
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3039 CAGGGGAGCCCCGGGTGCTCCCGTGTGCACACCTCAGACCTGGCTGCTTAGAGTAGCC 3098

Qy   3071 CGGCATGGTGGCTCATGTTCTCTTACCCTGAAAACTTTTAAAACTTGAAGTAGGTAGTAAGAT 3130
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3099 CGGCAATGGTGGCTCATGTTCTCTTACCCTGAAAACTTTAAAACTTGAAGTAGGTAGTAAGAT 3158

Qy   3131 GGCTTTCTTTTTTACCCTCCTGAGTGTTTATCACTCAGAAGTGATGGCTAAGGATACCAAAAAAAA 3190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3159 GGCTTTCTTTTTTACCCTCCTGAGTTTATCACTCAGAAGTGATGGCTAAGGATACCAAAAAAAA 3218

Qy   3191 CAAACAAAAAACAGAGAAACAAAAAAAAAACCTCACACAGCTCTCTTAGTACTCAGGT 3250
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3219 CAAACAAAAACAGAGAAACAAAAAAAAAACCTCAACAGCTCTCTTAGTACTCAGGT 3278

Qy   3251 TCATGCTGCAAAATCACTCGAGATTTTGTGTTTTTAGTAACCGTGCTTCCACATTTGCTGG 3310
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3279 TCATGCTGCAAAATCACTCGAGATTTTGTTTTTAAGTAACCGTGCTTCCACATTTGCTGG 3338

Qy   3311 AGGATGCTATTGTGAATGTGGGGCTCAGATGAGCAAGGGGCCAAAAAAAAAATTCCC 3370
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3339 AGGATGCTATTGTGAATGTGGGGCTCAGATGAGCAAGGTCAAAGGGGCCAAAAAAATTCCC 3398

Qy   3371 CCTCTCCCCCAGGAGTATTGAAAGATGAATGATGTTTATGGTTAAGTCTTCCTGGCACCTTC 3430
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3399 CCTCTCCCCCAGGAGTATTTGAAGATGAATGTTTATGGTTAAGTCTTCCTGGCACCTTC 3458

Qy   3431 CCCTTGCTTTGGTACAAAGGGGTGAAAGTCCTGTTGGTCTTGTTGTAGCATTTCCAGGATGATG 3490
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3459 CCCTTGCTTTGGTACAAAGGGGTATTTGAAGATAGATGTTTATGGTCATTTCCAGGATGATG 3518

Qy   3491 ATGTCAGCAGGGATGAGACAATCACCACCTTTAGGGCTTTTCCCTGGCAAGGGCCCATGTGGC 3550
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3519 ATGTCAGCAGGGATGACAATCACCACCTTTAGGGCTTTTCCCTGGCAGGGGCCCATGTGGC 3578
```

```
Qy   3551 TAGTCCTCACGAAGACTGGAGTAGGAGTAGAATGTTTGGGAGCTCAGGAAGGGTGGGTGGGAGTGGCC 3610
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3579 TAGTCCTCACGAAGACTGGAGTAGGAGTAGAATGTTTGGGAGCTCAGGAAGGGTGGGTGGGAGTGGCC 3638

Qy   3611 CTCTTCCAGGTGTGAGGGATAGGAGATAGGAAGGAGGAAGGAAGGTTAGGGAAAATCCATTCCCCACTCCCT 3670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3639 CTCTTCCAGGTGTGAGGGATAGGAGATAGGAAGGAGGAAGGAAGGTTAGGGAAAATCCATTCCCCACTCCCT 3698

Qy   3671 CTTTGGCCAAATGAGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAAACCCTAGTTCCTC 3730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3699 CTTTGGCCAAATGAGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAAACCCTAGTTCCTC 3758

Qy   3731 ATGAGAAGCTAGGAGCACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTTTCGGGGGACTA 3790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3759 ATGAGAAGCTAGGAGCAGCAGAACATCGTTCCCCTTATCTGAGCAGTGTTTCGGGGGACTA 3818

Qy   3791 CAGTGAAAAACCTTCTGGAGATGTTAAAAAGCTTTTTACCCACGATAGATGTTTTTTAA 3850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3819 CAGTGAAAAACCTTCTGGAGATGTTAAAAAGCTTTTTACCCACGATAGATTGTTGTTTTTAA 3878

Qy   3851 GGGGTGCTTTTTTTAGGGGCCATCACTGGAGATAAGAAAGTGCCATTTCAGAAAATGCCATC 3910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3879 GGGGTGCTTTTTTTAGGGGCATCACTGGAGATAAGAAAGTGCCATTTCAGAAAATGCCATC 3938

Qy   3911 GTAAATGGTTTTTAAACACCTTTTACCTAATTACAGGTGTCTATTTTATAGAAGCCAGAAAC 3970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3939 GTAAATGGTTTTTAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGAAGCAGACAAC 3998

Qy   3971 ACTTCTTTTTATGACTCTCAGACTTCTATTTTCATGTTACCAGTTTTTTTTTGTAACTCGCA 4030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3999 ACTTCTTTTTATGACTCTCAGACTTCTATTTTCATGTTACCATTTTTTTTTGTAACTCGCA 4058

Qy   4031 AGGTGTGGGGCTTTTGTAACTTCACAGGTGTGGGGAGGACTGCCTTGTTCTTGTTTCAACAGTTTG 4090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4059 AGGTGTGGGGCTTTTGTAACTTCACAGGTGTGGGGAGGACTGCCTTGTTCTTGTTTCAACAGTTTG 4118

Qy   4091 TCTCCACTCGGTTTCTACTTCACAGGTGCAAAAGATGACAGAGATGCCCCAGAGTTTTAACCTTTTCT 4150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4119 TCTCCACTCGGTTTCTACTTCACAGGTGCAAAAGATGACAGAGATGCCCCAGAGTTTTACCTTTTTCT 4178

Qy   4151 GGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAAA 4195
          |||||||||||||||||||||||||||||||||||||||||
Db   4179 GGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAAAAA 4223
```

RESULT 6
AF151974
LOCUS   AF151974   4163 bp   mRNA   linear   ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 6
            (Dnmt3b) mRNA, complete cds.
ACCESSION   AF151974
VERSION   AF151974.1 GI:8347130
KEYWORDS   .
SOURCE   Mus musculus
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4163)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE   Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 4163)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China
FEATURES   Location/Qualifiers
     source   1..4163

```
                    /organism="Mus musculus"
                    /strain="KM"
                    /db_xref="taxon:10090"
                    /dev_stage="8-9 day old embryo"
gene            1..4163
                    /gene="Dnmt3b"
CDS             297..2816
                    /gene="Dnmt3b"
                    /function="methylates cytosine in DNA"
                    /note="alternatively spliced"
                    /codon_start=1
                    /product="DNA cytosine-specific methyltransferase isoform
                    6"
                    /protein_id="AAF74520.1"
                    /db_xref="GI:8347131"
                    /translation="MKGDSRHLNEEEGASGYEECIIVNGNPESDQSSDPTKDAPSPVLE
                    AICTEPVCTPETRGCRRSSSRLSKREVSSLLNYTQDMTGDGRDDEVDDGNGSDILMPK
                    LTRETKDTRTRSESPAVRTRHSNGTISSLERQRASPRITRGRGQRHHVHVQEYEVEFPATR
                    SRRRASSSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDQEGMDTTQVDAESLY
                    GDSTEYQDDKEPGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                    SEISADKLVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKFFSSSPGESLED
                    QLKPMLEWAHGGFKPTGIEGLKPNKKQPENKSRRTTNDSAASESPPPKRLKTNSYGG
                    KDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVSFHPLFEGGLCQSCRDRFLEL
                    FYMYDEDGYQSYVCVCCGRELLLCSNTSCCRCFCVECLEVLVGAGTAEDAKLQEPWS
                    CYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPDLEEFEPPKLYPAIPAAKRRPIRVLSL
                    FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYVNDVRKITKKNIEE
                    WGPFDLVIGGSPCNDLSNVNPARKGLYEGTGIRGRLFFEFYHLLNYTRPKEGDNRPFFWMF
                    ENVVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARYFWGNLPGMNRPVWASKNDKL
                    ELQDCLEPSRTAKLKKVQTITTKSNSIRQGKNQLFPVVMNGRDDVLMCTELERIFGFP
                    AHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"

BASE COUNT    1032 a   1074 c   1128 g    929 t
ORIGIN

Query Match       90.9%;  Score 3812.8;  DB 10;  Length 4163;
Best Local Similarity   98.3%;  Pred. No. 0;
Matches 3878;  Conservative   0;  Mismatches   7;  Indels  60;  Gaps   1;

Qy    251 GCCTTCTTTCAGGAAACAANTGAAGGGAGACAGCAGCAGACAGACATCTGAATGAATGAAGAGAGAGGGTGCC 310
          |||||
Db    279 GCCTCACGACAGGAAANCAAATGAAGGGAGACAGCAGCACAGACATCTGAATGACAANGAGAGGGTGCC 338

Qy    311 AGCGGGTATAGAGGAGTGCAATTATCGTTAATGGGAGAACTTCAGTGACCAGTGCCTCAGACACG 370
          |||||||||||||||||||
Db    339 AGCGGGTATAGAGGAGTGCATTATATCGTTAATGGGAACTTCAGTGACCAGTGCCTCAGACACG 398

Qy    371 AAGGATGGCTCCCTCACCCCCCAGTGTTGGAGGGCAAATCTGCACAAGGCCAGTCTGCACACCA 430
          |||||||||||||||||||||
Db    399 AAGGATCTCCCTCACCCCCAGTCTTGGAGGGCAAATCTGCACAAGGCCAGTCTGCACACCA 458

Qy    431 GAGACCAGAGGCCGCAGGTCAAGCTCCCCGGCTGTCTAAGAGGGAGGGTCTCCCAGCCTTCTG 490
          |||||||||||||||||||||
Db    459 GAGACCAGAGCCCGCAGGTCAAGCTCCCCGGCTGTCTAAGAGGGAGGGTCTCCCAGCCTTCTG 518

Qy    491 AATTACACGCAGGACATGACAGGAGCAGATGAGACAGAGAGAGTAGAATGCATGGGAAT 550
          ||||||||
Db    519 AATTACACGCAGGACATGACAGGAGATGGAGACACAGAGATGAATGAAGTAGATGGGAAT 578

Qy    551 GGCTCTGATATTCTAATGCCAAAGCTCACCCCGTGACCAAGGACACCAGGACGCGCGTCT 610
          |||||||||||
Db    579 GGCTCTGATATTCTAATGCCAAAGCTCACCCCGTGAGACCAAGGACACCAGGACGCGCGTCT 638

Qy    611 GAAAGCCCGGCCTGTGTCCGGAACCCGACACATAGCAAATGGGACCAATAGCCAAAGA 670
          |||||||||||
Db    639 GAAAGCCCGGCCTGTCCGGAACCCGACATAGCAATGGGACCAATAGCCAAAGA 698

Qy    671 GCCTCCCCAGAAATCACCCCAGGGTGGGCAGGGCGCGCCACCATGTCCAGGAGTACCCGTGTG 730
          |||||||||||
Db    699 GCCTCCCCCAGAAATCACCCCGAGGTGCGGCAGGGCGCGCCACCATGTCCAGGAGTACCCGTG 758

Qy    731 GAGTTTCCGGCTACCAGGGTCTCGGAGGACCGTCGACCAGTCGCAAGCACGCACCATGG 790
          |||||||||||
Db    759 GAGTTTCCGGCTACCAGGGTCTCGGAGGACCGTCGACCAGTCGCAAGCACGCACCATGG 818

Qy    791 TCATCCCCTGCCAGCCGTCGACTTCATGGAAGAAGAAGTGACACCTAAGAGGCCGTCAGTACCCCA 850
          ||||||
Db    819 TCATCCCCTGCCAGCCGTCGACTTCATGGAAGAAGAAGTGACACCTAAGAGGCGTGGATACCCCA 878

Qy    851 TCAGTTGACTTGAGCCACGGATGGAGATCAGGAGGGTATGGATACCACACAGGTGGATGCA 910
          |||||||||||
Db    879 TCAGTTGACTTGAGCCACGGATGGAGATCAGGAGGGTATGGATACCACACAGGTGGATGCA 938

Qy    911 GAGAGCAGAGATGGAGACAGCACAGAGTATCAGGATGATGAAAGAGTTGGAATAGGTGAC 970
          |||||||||||
Db    939 GAGAGCAGATATATGGGAGAGACAGACAGAGTATCAGGATGATGAAAGAGTTTGGAATAGGTGAC 998

Qy    971 CTCGTGTGGGGGAAAGATCAAGGGCCTTCTCCCTGGTCCTGCCATGGGTGGGTCCTGGAAA 1030
          |||||||||||
Db    999 CTCGTGTGGGGGAAAGATCAAGGGCCTTCTCCCTGGTCCTGCCATGGGTGGGTCCTGGAAA 1058

Qy   1031 GCCACCTCCAAGGCGACAGGCCCATGCCCGGAAGTGCCGCTGGGGTACAGTGGTTGGGTGAGTGGC 1090
          |||||||||||
Db   1059 GCCACCTCCAAGCGACAGGCCCATGCCCGGAAATGCCGCTGGGGTACAGTGGTTGGGTGATGGC 1118

Qy   1091 AAGTTTTCTGAGATCTCTGCTGACAAAACTGGTTGGCTTCTGGGGCGTGTTCAGCCAGCACTTT 1150
          |||||||||||
Db   1119 AAGTTTTCTGAGATCTCTGCTGACAAAACTGGTGGCTTCTGGGGCGTGTTCAGCCAGCACTTT 1178

Qy   1151 AATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTACCACACTCTTGGAG 1210
          |||||||||||
Db   1179 AATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTACCACACTCTTGGAG 1238

Qy   1211 AAAGCCAGGGTTCGAGCGTGGCCAAGAACCTTCTCCAGCAGTCCTGGAGAGTCCAGCACTT 1270
          |||||||||||
Db   1239 AAAGCCAGGGTTCGAGCGTGGCCAAGAACCTTCTCCAGCAGTCCTGGAGAGTCACTGGAGGAC 1298

Qy   1271 CAGCTGAAGCCCCATGCTTGGGAGTGGGCCCCACGGTGGCCTTCAAGCCTACTGGGATCGGAGGGC 1330
          |||||||||||
Db   1299 CACCTGAAGCCCCATGCTTGGGAGTGGGCCCCACGGTGGCCTTCAAGCCTACTGGGATCGAGGGC 1358

Qy   1331 CTCAAAACCCAACAAGAAGCAACCCAGTGGTTAATAAGCTCGAAGGTGCCGTCGTTCAGACAGT 1390
          |||||||||||
Db   1359 CTCAAAACCCAACAAGAAGCAACC 1381

Qy   1391 AGGAAACTTAGAACCCAGGAGACGCGACAGACAAAAACTCGAAGACACCACCAATGACACTCT 1450
          |||||||||||
Db   1382 ------------------------CGAGAACAAAAGTCGAAGACGCCACCAATGACACTCT 1418

Qy   1451 GCTGCTTCTGACTCCCCCCACCCAAGCGCCTCAGACACAAATAGCTATGCGGGAAGGAC 1510
          |||||||||||
Db   1419 GCTGCTTCTGAGTCCCCCCACCCAAGCGCCTCAACAGCAAATAGCTATGCGGGAAGGAC 1478

Qy   1511 CGAGGGGAGGGATGAGGAGAGAGCCGAGAACGGAGTCCTTCTCAAAGTCACCAACAACAAGGGC 1570
          |||||||||||
Db   1479 CGAGGGGAGGGATGAGGAGAGAGCCGAGAACGGATGGCTTCTCAAAGTCACCAACAACAAGGGC 1538

Qy   1571 AATCTGGAAGACCGCGTCTTTGTCCCTGTGGAAAGAAGAGAACAAAGCCCTGTGTCCTTCCACCCCCTC 1630
          |||||||||||
Db   1539 AATCTGGAAGACCGCGTCTTTGTCCCTGTGGAAAGAAGAAACCCTGTGTCCTTCCACCCCCTC 1598

Qy   1631 TTTGAGGGTGGGCCTCTGTCTCAGAGTTGCCCGGGAATCCCTTCCTAGAGACCTCTTCTACATGTAT 1690
          |||||||||||
Db   1599 TTTGAGGGTGGGCCTCTGTCTCAGAGTTGCCCGGGAATCCCTTCCTAGAGACCTCTTCTACATGTAT 1658

Qy   1691 GATGAGGACGGCTATACAGTCCTTACTGCCACCGTGTGCTGTGCAGGGCCTGAAACTGCTGCTG 1750
          |||||||||||
Db   1659 GATGAGGACGGCTATACAGTCCTTACTGCACCGTGTGCTGTGCAGGGCCTGAAACTGCTG 1718

Qy   1751 TGCAGTRAACACAAAGCTGCTGCAGATGCCTTCTTGTGTCTGGAGTGTGTCTGGAGGTGGGGC 1810
          |||||||||||
Db   1719 TGCAGTRAACACAAAGCTGCTGCAGATGCCTTCTGTGTCTGGAGTGTGTCTGGAGGTGGGGC 1778

Qy   1811 GCAGGCACAGCTGAGGATGCCAAGCTGCAGGAAACCCTGGAGCTGCTGTATATGTGCCTCCT 1870
          |||||||||||
Db   1779 GCAGGCACAGCTGAGGATGCCAAGCTGCAGGAAACCCTGAGCTGCTGTATATGTGCCTCCCT 1838

Qy   1871 CAGCGCGTGCCATGGGGTCCTCCGACGCCGAGGCAGGAGAAAAGATTGGAACAATGCGCGCCTGCAAGACTTC 1930
```

```
Qy   4091 TCTCCACTGGTTTCTTAATTTTTTAGGTGTCCAAAGATGACAGATGCCCAGAGTTTACCTTTCT 4150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4059 TCTCCACTGGTTTCTTAATTTTTTAGGTGTCCAAAGATGACAGATGCCCAGAGTTTACCTTTCT 4118

Qy   4151 GGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAA 4195
          |||||||||||||||||||||||||||||||||||||| ||||||
Db   4119 GGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAA 4163
```

RESULT 7
LOCUS   AY078427   4006 bp   mRNA   linear   ROD 01-JUL-2002
DEFINITION   Mus musculus DNA cytosine methyltransferase 3b6 (Dnmt3b6) mRNA,
             complete cds; alternatively spliced.
ACCESSION   AY078427
VERSION   AY078427.1 GI:21655120
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4006)
  AUTHORS   Chen,T., Ueda,Y. and Li,E.
  TITLE   Dnmt3a2 encoded by transcripts from an intronic promoter of Dnmt3a
             is expressed abundantly in ES cells and germ cells
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 4006)
  AUTHORS   Chen,T., Ueda,Y. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Cardiovascular Research Center,
             Massachusetts General Hospital, 149 13th Street, Charlestown, MA
             02129, USA
FEATURES   Location/Qualifiers
     source   1..4006
                     /organism="Mus musculus"
                     /strain="129S4/SvJae"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="84.0 cM"
     gene   1..4006
                     /gene="Dnmt3b6"
     CDS   269..2659
                     /gene="Dnmt3b6"
                     /function="de novo methylation"
                     /note="contains PHD and PWWP domains; alternatively
                     spliced"
                     /codon_start=1
                     /product="DNA cytosine methyltransferase 3b6"
                     /protein_id="AAL85481.1"
                     /db_xref="GI:21655121"
                     /translation="MKGDSRHLNEEEGASGYEECIIVNGNFSDQSSDTKDAPSPVLE
                     AICTEPVCTPETRGRRSSSRLSKREVSSLLNTTQDMTGDGDRDDEVDDGNGSDILMPK
                     LTRETKDTRTRSESPAYRTRHSNGTSSLERQRASPRITRGRQGRHHVQEPVEFPATR
                     SRRRAASSSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDCDQEGMDTQVDAESRD
                     GDSTEYQDDKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                     SEISADKIVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTFSSSPGESLED
                     QLKPMLEWAHGGEKPTGIEGLKPNKQPVVNKSKVRRSDSRNLEPRRENKSRRRTTN
                     FHPLFEGGLCQSCRDRFELELFYMIDEDGYQSYCTVCCEGRELLLCSNTSCCRCFVEC
                     DSAASESPPPKRLKTNSYGGKDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVS
                     LEVLVGAGTAEDAKLQEFWSCTMCLPQRCHGVLRRRRDWNMRLQDFFTTDPDLEEFEP
                     PKIYPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKH
                     EGQIKVNDVRKITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFY
                     HLLNYTRPKEGDNRPFFWMFENVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARY
                     FWGNLPGMNRIFGFPAHYTDVSNMGRGAROKLLGRSWSVPVIRHLFAPLKDYFACE"

BASE COUNT   1001 a   1017 c   1075 g   913 t
ORIGIN
```

```
Qy      1 GAATTCCGGGCGCGCCGGGGTTAAGCGGCCCAAGTAAACGTAGCCGCAGCCAATCGGCCGCGGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAATTCCGGGCGCGCCGGGGTTAAGCGGCCCAAGTAAACGTAGCCGCAGCCAATCGGCCGCGGA 60
```

(GenBank sequence alignment — Qy / Db paired rows continue with full identity matching through position 1140)

```
Query Match     90.8%;  Score 3807;  DB 10;  Length 4006;
Best Local Similarity   95.5%;  Pred. No. 0;
Matches 4006;  Conservative 0;  Mismatches 0;  Indels 189;  Gaps 1;
```

REFERENCE　　1　(bases 1 to 4149)
AUTHORS　　Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
　　　　　　Shen,Y.
TITLE　　Cloning of full-length Dnmt3b cDNA and its alternative splicing
　　　　　　isoforms in mouse embryonic tissue
JOURNAL　　Unpublished
REFERENCE　　2　(bases 1 to 4149)
AUTHORS　　Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
　　　　　　Shen,Y.
TITLE　　Direct Submission
JOURNAL　　Submitted (17-MAY-1999) Department of Biochemistry and Molecular
　　　　　　Biology, Institution of Basic Medical Sciences, Chinese Academy of
　　　　　　Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
　　　　　　P.R.China

FEATURES　　　　Location/Qualifiers
source　　　　1..4149
　　　　　　/organism="Mus musculus"
　　　　　　/strain="KM"
　　　　　　/db_xref="taxon:10090"
　　　　　　/dev_stage="8-9 day old embryo"

gene　　　　1..4149
　　　　　　/gene="Dnmt3b"

CDS　　　　412..2802
　　　　　　/gene="Dnmt3b"
　　　　　　/function="methylates cytosine in DNA"
　　　　　　/note="alternatively spliced"
　　　　　　/codon_start=1
　　　　　　/product="DNA cytosine-specific methyltransferase isoform
　　　　　　4"
　　　　　　/protein_id="AAF74518.1"
　　　　　　/db_xref="GI:8347126"
　　　　　　/translation="MKGDSRHLNEEEGASGYEECIIVNGNFSDQSSDTKDAPSPPVLE
　　　　　　AICTEPVCTPETRGRRRSSSRLSKREVSSLLNYTODMTGDGDRDDEVDDGNSDILMPK
　　　　　　LTREKDTRTRSESPAVRTRHSNGTSSLERQRASPRITRGRQORHHVOEYPVEFPATR
　　　　　　SRRRRASSSSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDQOEGMDTTQVDAEESIY
　　　　　　GDSTEYODDKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKROAMPGMRWVQWFGDKHF
　　　　　　SEISADKLVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTFSSSPGESLED
　　　　　　QLKPMLEWAHGGFKPTGIEGLKPNKKQPVVNKSKVRRSDSRNLEPRRRENKSRRRTN
　　　　　　DSAASESPPPKRLKTNSYGGKDRQEDEESRERNASEVTNNKGNLEDRCLSCGKKNPVS
　　　　　　FHPLFEGGLCQSCRDRFLELFYMYDEDGYQSYCTVCCEGRELLLCSNSCCRCFCVEC
　　　　　　LEVLVGAGTAEDAKLQEPWSCYMCLPQRCHGVLRRRKDWNMRLQDFTTDPLEEEFP
　　　　　　PKLYPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKH
　　　　　　EGOIKYVNDVRKITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEY
　　　　　　HLLNYTRPKEGDNRPFWWMFENNVVAMKVNDKKDISRFLACNPVMWIDAIKVSAAHRARY
　　　　　　FWGNLPGMNRIFGFPAHYIDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"

BASE COUNT　　1013 a　　1123 g　　931 t
ORIGIN　　　1082 c

Query Match　　90.58%;　　Score 3794.8;　　DB 10;　　Length 4149;
Best Local Similarity　　95.4%;　　Pred. No. 0;
Matches 3995;　　Conservative　　0;　　Mismatches　　2;　　Indels 189;　　Gaps　　1;

Qy　　10　GCGCGGCGGGTTAAGCGGCGGCCCAAGTAAAACGTAGCGCAGCGATCGGCGCCCGGAGATTCGCGA　　69
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　153　GCCGCGGGGGTTAAGGCGGCCCAAGTAAAACGTAGCGCAGCGA**TCGGCGCCCGGAGATTCGCGA**　212

Qy　　70　ACCCGACAACTCCGCGCGCCGCGCGCCGCCGCCAGGAACCCGCGGGCGGCGGCGCGCGCCGCGCTA　129
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　213　ACCCGACAACTCCGCGCCGCCGCGCCGCCGCCAGGACCCGCGGGCGGCGGCGCGCCGGCGCGCTA　272

Qy　　130　CAGCCAGGCCTCACGACACAGGCCCCGCGCTGAGGCCTTGTGCCAGACCTTGGAAACCTCAGGTATA　189
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　273　CAGCCAGGCCTCACGACACAGGCCCCGCGCTGAGGCCTTGTGCCAGACCTTGGAAACCTCAGGTATA　332

Qy　　190　TACCTTTCCAGACGCGGGGATCTCCCCTCCCCCATCCATAGTGCCTTGGGACCAAAATCCAG　249
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　333　TACCTTTCCAGACGCGGGATCTCCCCTCCCCCATCCATAGTGCCTTGGGACCAAAATCCAG　392

Qy　　250　GGCCTTCTTTCAGGAAAACAATGAAAGGGAGACAGCAGACAATCTGAATGAAGAAGAGAGGGTGC　309
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　393　GGCCTTCTTTCAGGAAAACAATGAAAGGGAGACAGCAGACAATCTGAATGAAGAAGAGAGGGTGC　452

Qy　　310　CAGCGGGTATGAGGAGTGCATTATCGTTAATGGGAACTTCAGTGACCAGTCCTCCAGACAC　369
Db

```
Qy  2530 TTCAAAGAATGATAAGCTCGAGCTGCCTGGAGTTCAGTAGGAGGACAGCAAAGTT 2589
Db       ------------------------------------------------------- 2658
Qy  2590 AAAGAAAGTGCAGACAATAACCACCAAGTCGAACTCATCAGACAGGGCAAAAACCACCT 2649
Db  2659 ---------------------------------------------------------- 2658
Qy  2650 TTTCCCTGTAGTCATGAATGGCAAGGACGACGCGTTTGTGGTGCACTCGAAAGGAT 2709
Db  2659 ----------------------------------------------------AGGAT 2663
Qy  2710 CTTCGGCTTCCCTGCTCACTACACGGACGTGTCCAACATGGGCCGCGGCGCCGGTCAGAA 2769
Db  2664 CTTCGGCTTCCCTGCTCACTACACGGACGTGTCCAACATGGGCCGCGGCGCCGGTCAGAA 2723
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2770 GCTGCTGGGCAGGTCCTGGAGTGTACGGTCATCAGACACACCTGTTTGCCCCCTTGAAGGA 2829
Db  2724 GCTGCTGGGCAGGTCCTGGAGTGTACGGTCATCAGACACACCTGTTTGCCCCCTTGAAGGA 2783
Qy  2830 CTACTTTGCCTGCTGAATAGTTCTACCCAGAGACTGGGAGCTCTCGGTCAGAGCCCAGTCC 2889
Db  2784 CTACTTTGCCCGTGTGAATAGTTCTACCCAGGACTGGGAGCTCTCGGTCAGAGCCCAGTGCC 2843
Qy  2890 CAGAGTCACCCCTCCCTGAAGGCACCTCACCTGTCCCCCTTTTTAGCTCACCTGTGTGGGG 2949
Db  2844 CAGAGTCACCCCTCCCTGAAGGCACCTCACCTGTCCCCCTTTTTAGCTCACCTGTGTGGGG 2903
Qy  2950 CCTCACATCACTGTACCTCACCTTCTCCTGCTCAGTGGGAGCCAGAGCCTCCTGGCCCTT 3009
Db  2904 CCTCACATCACTGTACCTCACCTTCTCCTGCTCAGTGGGAGCCAGAGCCTCCTGGCCCTT 2963
Qy  3010 GCAGGGGAGCCCCCGGTGCTCCCTCCGTGTGCACAGCTCAGACCTGGCTTAGAGTAGC 3069
Db  2964 GCAGGGGAGCCCCCGGTGCTCCCTCCGTGTGCACAGCTCAGACCTGGCTTAGAGTAGC 3023
Qy  3070 CCGGCAATGGTGCTCATGTTCTCTTACCCTGAAAACTTTAAAACTTGAAGTAGGTAGTAAGA 3129
Db  3024 CCGGCAATGGTGCTCATGTTCTCTTACCCTGAAAACTTTAAAACTTGAAGTAGGTAGTAAGA 3083
Qy  3130 TGGCTTTCTTTTACCCTCCTCAGTTTTTATCACTCAGAGATGATGGCTAAGATACCAAAAAA 3189
Db  3084 TGGCTTTCTTTTACCCTCCTGAGTTTATCACTCAGAGATGATGGCTAAGATACCAAAAAAA 3143
Qy  3190 ACAAACAAAAACAGAAACAAAACAAAAAAACCTCAACAGCTCTCTTTAGTACTCACTCAGG 3249
Db  3144 ACAAACAAAAACAGAAACAAAACAAAAAAACCTCAACAGCTCTCTTTAGTACTCACTCAGG 3203
Qy  3250 TTCAATGCTGCAAAAATCACTTTGAGATTTGTTTTTGTTTTAAGTAACCCGTGCTCTCCACATTTGCTG 3309
Db  3204 TTCAATGCTGCAAAAATCACTTTGAGATTTGTTTTTGTTTTAAGTAACCCGTGCTCTCCACATTTGCTG 3263
Qy  3310 GAGGATGCTATTGTGAATGTGGGCTCAGATGAGCAAGGTCAAAGGGGCCAAAAAAATTCC 3369
Db  3264 GAGGATGCTATTGTGAATGTGGGCTCAGATGAGCAAGGTCAAAGGGGCCAAAAAAATTCC 3323
Qy  3370 CCCTCTCCCCCAGGAGTATTTGAAAGATGAATGATGTTTATGGTTTAAGTCTTCCTGGCACCTT 3429
Db  3324 CCCTCTCCCCCAGGAGTATTTGAAAGATGAATGATGTTTATGGTTTAAGTCTTCCTGGCACCTT 3383
Qy  3430 CCCCTTGCTTTGGTACAAGGGCTGAAGTCCTGTTGGGTCTTGTTGTAGCATTTCCCAGGATGAT 3489
Db  3384 CCCCTTGCTTTGGTACAAGGGCTGAAGTCCTGTTGGGTCTTGTTGTAGCATTTCCCAGGATGAT 3443
Qy  3490 GATGTCAGCAGGGATGACATCACCACCACCTTTAGGGCCTTTTCCCTGGCCAGGGGGCCCATGTGG 3549
Db  3444 GATGTCAGCAGGGATGACATCACCACCACCTTTAGGGCCTTTTCCCTGGCCAGGGGGCCCATGTGG 3503
Qy  3550 CTAGTCCTCACGAAGAACTCGGAGTAGAAATGTTTGGAGCTCAGGAAGGGGTGGGGTGGAGTGGC 3609
Db  3504 CTAGTCCTCACGAAGAACTCGGAGTAGAAATGTTTGGAGCTCAGGAAGGGGTGGGGTGGAGTGGC 3563
```

```
Qy  3610 CCTCTCCCCAGGTGTGTGAGGGGTATACCGAAGGAGGAGGAGGAGGAGCCTTAGGGAAAATCCAATTCCCCACTCCC 3669
Db  3564 CCTCTCCCCAGGTGTGTGAGGGGTATACCGAAGGAGGAGGAGGAGGAGCCTTAGGGAAAATCCAATTCCCCACTCCC 3623
Qy  3670 TCTTGCCAAATGAGGGGCGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCCTTAGTTCCT 3729
Db  3624 TCTTGCCAAATGAGGGGCGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCCTTAGTTCCT 3683
Qy  3730 CATGAGAAGCTAGGACCAGAAGCACACATCGTTCCCCTTATCTGACCAGTGTGTTTGGGGAACT 3789
Db  3684 CATGAGAAGCTAGGACCAGAAGCACACATCGTTCCCCTTATCTGACCAGTGTGTTTGGGGAACT 3743
Qy  3790 ACAGTGAAAAACCTTCTGGGAGATGTTAAAAAGCGTTTTTACCCCACGATAGATTGTGTTTTTA 3849
Db  3744 ACAGTGAAAAACCTTCTGGGAGATGTTAAAAAGCGTTTTTACCCCACGATAGATTGTGTTTTTA 3803
Qy  3850 AGGGGTGCCTTTTTTTTAGGGGGCATCACTGGAGATAAGAAAGCTGCATTTCAGAAAATGCCAT 3909
Db  3804 AGGGGTGCCTTTTTTTTAGGGGGCATCACTGGAGATAAGAAAGCTGCATTTCAGAAAATGCCAT 3863
Qy  3910 CGTAAATGGTTTTTAAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGAAGGCAGACAA 3969
Db  3864 CGTAAATGGTTTTTAAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGAAACCAGACAA 3923
Qy  3970 CACTTCTTTTTATGACTCTCAGACTTCTATTTTCATGTTACCATTTTTTTTGTAACTCGC 4029
Db  3924 CACTTCTTTTTATGACTCTCAGACTTCTATTTTCATGTTACCATTTTTTTTGTAACTCGC 3983
Qy  4030 AAGGTGTGGGCGCTTTTGTAACTTCACAGGTGTGGGGGAGGAGCTGCCTTGTTTCAACAGTTT 4089
Db  3984 AAGGTGTGGGCGCTTTTGTAACTTCACAGGTGTGGGGGAGGAGCTGCCTTGTTTCAACAGTTT 4043
Qy  4090 GTCTCCACTGGTTTCTAAATTTTTAGGTGCAAAGATGACAGATGCCCAGAGTTTACCTTTC 4149
Db  4044 GTCTCCACTGGTTTCTAAATTTTTAGGTGCAAAGATGACAGATGCCCAGAGTTTACCTTTC 4103
Qy  4150 TGGTTGATTAAAGTTGTATTTCTTAAAAAAAAAAAAAAAAAAAA 4195
Db  4104 TGGTTGATTAAAGTTGTATTTCTTAAAAAAAAAAAAAAAAAAAA 4149
```

RESULT 9

AF068628

LOCUS   AF068628   3946 bp   mRNA   linear   ROD 06-DEC-1999
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B3 (Dnmt3b) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF068628
VERSION   AF068628.2 GI:6449473
KEYWORDS   .
SOURCE   Mus musculus.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 3946)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE   Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
  JOURNAL   Nat. Genet. 19 (3), 219-220 (1998)
  MEDLINE   98324766
  PUBMED   9662389
REFERENCE   2 (bases 1 to 3946)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REFERENCE   3 (bases 1 to 3946)
  AUTHORS   Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
  REMARK   Sequence update by submitter
COMMENT   On Nov 18, 1999 this sequence version replaced gi:3327983.
FEATURES   Location/Qualifiers

```
source          1..3946
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="2"
                /map="84.0 cM"
gene            1..3946
                /gene="Dnmt3b"
gene            1..3946
                /gene="Dnmt3b"
misc_feature    1..3946
                /gene="Dnmt3b"
                /note="similar to EST sequences deposited in GenBank
                Accession Numbers AA116694, AA119979, AA177277, AA210568,
                AA407106, and AA57617"
CDS             269..2599
                /gene="Dnmt3b"
                /note="de novo DNA methylation"
                /function="de novo DNA methylation"
                /note="alternatively spliced product; contains Cys-rich
                region"
                /codon_start=1
                /product="DNA cytosine-5 methyltransferase 3B3"
                /protein_id="AAC40180.2"
                /db_xref="GI:6449474"
                /translation="MKGDSRHLNEEGASGYEECIIVNGNFSDQSSDTKDAPSPVLE
                AICTEPVCTPETRGRRSSRLSKREVSSLLNTYQDMTCDGDRDDEVDDGNGSDIIMPK
                LTRETKDTRTRSSESPAVRTRHRSNGTSSLERQRASPRITRGRQGRHHVQEYPVEPATR
                SRRRASSASSTPWNSSPASVDFMEEVTPKSVSTPSVDLSQDGDGEGMDTTVVDAESRD
                GDSTEYQDDKEFGIGDLVWGKIKGPSEWPAMVVSWKATSKRQAMPGMRNVQWFGDGKF
                SEISADKLVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTFSSSSPGESLED
                QLKPMLEWAHGGRKPTGIEGLKPNKKQPENKSRRRTTNDSAAESPPPKRLKTNSYGG
                KDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVSFHPLFEGGLCQOSCRDRFLEL
                FYMYDEDGYQSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGAGTAEDAKLQEPWS
                CYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPDLEEEPEKPKLPEAIPAAKRRPIRVLSL
                FDGIATGVLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKVNDVRKITKNIEE
                WGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYHLLNYTRPKEGDNRPFWWMF
                ENVVAMKVNDKDIISRFLACNPVMIDAIKVSAAHRARYFWGNLPGMNRIFGFPAHYTD
                VSNMGRGARQKLLGRSWSWSVPVIRHLFAPLKDYFACE"
BASE COUNT      983 a   1006 c   1056 g    901 t
ORIGIN
```

```
Query Match     87.7%;  Score 3677;  DB 10;  Length 3946;
Best Local Similarity   94.1%;  Pred. No. 0;
Matches 3946;  Conservative 0;  Mismatches 0;  Indels 249;  Gaps 2;
```

```
Qy      1 GAATTCCGGCCGGCCGGGGTTAAGGCCAAGTAAAACGTAGCGCAACGCAGCGATCGGCGGCCGGGA   60
          ||||||||||| ||||||||||||| ||| ||||||||||||||| |||||||||||| | |||||
Db      1 GAATTCCGGCCGCGCCGGGGTTAAGCGGCCCAAGTAAAACGTAGCGCAGCGATCGGCGCGCGGGA   60

Qy     61 GATTCGGCGAACCCGACACTCCGCGCCGCCGCCCGCGCCGCGCCCAGGACCCCGCGCCGCGGGC  120
          || |||||||||||| |||| |||||||||||||||||||| ||||||||| |||| ||||||||
Db     61 GATCGGCGAACCCGACACTCCGCGCCGCCGCCGCGCCGCGCCCAGGACCCCGCGCCGCGGGC    120

Qy    121 GCCGCGCGCTACAGCCCAGCCTCACGACACAGCGGCCCGCTGAGCCGGTTGGAAACC           180
          ||||||||| ||||||||||||||||||||||| || |||||||||||||||||||||
Db    121 GCCGCGCTACAGCCCACGACACAGGCGCCCCGCCTGAGCGCTTGTGCCCAGACCTTGGAAACC    180

Qy    181 TCAGGTATATACCTTTCCAGACGCGGGGATCTCCCCGCTCCCCGCATCCATCCAGTCCTTGGGAC  240
          ||||||||||||||||||||||||||||||| || ||| |||||||||||||||||||||||||
Db    181 TCAGGTATATACCTTTCCAGACGCGGGGATCTCCCCATCCCAGTGCCCTAGTGCCTTGGGAC      240

Qy    241 CAAATCCAGGGCCTTCTTTCTTCAGGAAAACAAATGAAGGGAGACAGCAGACATCTGAATGAAGA  300
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAATCCAGGGCCTTCTTTCTTCAGGAAAACAAATGAAGGGAGACAGCAGACATCTGAATGAAGA  300

Qy    301 AGAGGGTGCCAGCGGGGTATGAGGAGTGTCCATTTATTCGTTAATTGGGAACTTCAGTGACCAGTC  360
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    301 AGAGGGTGCCAGCGGGGTATGAGGAGTGTGAGGCATTATCGTTAATTGGGAACTTCAGTGACCAGTC 360

Qy    361 CTCAGACACGAAAGGATGCTCCCTCACCCCCAGTCTTGAGGCAAATCTGCACAGAGCCAGT        420
          ||| ||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||
Db    361 CTCAGACACGAAAGGATGCTCCCTCACCCCAGTCTTGGAGGCAAATCTGCACAGAGCCAGT       420

Qy    421 CTGCACCAGCAGAGACCAGAGGCCGCGCAGGTCAAGCTCCCGGCGTGTCTAAGCAGCACGTCTC   480
          || ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGGACCAGCAGAGACCAGAGGCCGCGCAGGTCAAGCTCCCGGCGTGTCTAAGCAGGAGGTCTC    480
```

```
Qy    481 CAGCCTTCTCGAATTACACGCAGGACAATGACAGGCAGAGGAGAGATGGCAGGAGAGAGATGAGA  540
          || |||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGGCCTTCTGGAATTACACGCAGGACATGACAGGGAGGAGAGATGGCAGGAGAGATGGAGA     540

Qy    541 TGATGGGAATGGGCTCTGATATTCTAATGGCAAAAGCTCACCCGGTCAGACAACACCAG        600
          ||||||||||||| || |||||||||||||||||||||||||||||||||||||| ||||
Db    541 TGATGGGAATGGGCTGTGATATTCTAATGCCAAAAGCTCACCGTGAAACCAAGGACACACAG     600

Qy    601 GACGCGCTCTGAAAGCCCCGGCGTGTCCGAAACCCGACATAGCCAATGGGACCTTCGAGGA      660
          || ||||||||||| ||||||||||||||||||||||| ||||||||||||||||||||||
Db    601 GAGGCAAAGAGCCTCCCCAGAATCACCCGAGGTCGCGGCATCTCCTCAGGA             660

Qy    661 GAGGCAAAGAGCCTCCCCCAGAATCACCCGAGGTCGCGGCATCTCTTCAGCAAG            720
          ||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGGCAAAGAGCCTCCCCCAGAATCACCCGAGGTCGCGGCATCTCTTCAGCAAG            720

Qy    721 GTACCCTGTGTGGAGTTTCCGGGTACCAGTCTTCGGAGACGTGACCATCTCTTTCAGCAAG       780
          || |||| ||||||||||||| |||| |||| | ||||||||||||||||||||||||||
Db    721 GTACCCTGTGTGGAGTTTTCCGGGCTACCAGTCTTCGGAGACGTGACCATGTGCCACAAG       780

Qy    781 CACGCCCATGGTCATCCCCTGCCAGCGTCGACTTCATGGAAGAAGAGTGACACTTAAGAGCGT    840
          |||| ||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db    781 CACGCCCATGGTCATCCCCTGCCAGCGTCGACTTCATGGGAGAAGAGTGACACTTAAGAGCGT    840

Qy    841 CAGTACCCCATCAGTTGACTTGAGCCACAGGATGGAGATGAAGGTGTATGGATAACACACA    900
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CAGTACCCCATCAGTTGACTTGAGCCAGAGGATGAGGAGGGTATGGATACCACACA       900

Qy    901 GGTGGATGCAGAGACGAGAGAGAGAGACAGACAGAGTATCAGGATGATGAAAGAAGTTTGG     960
          |||||| ||||||||||||||||||||| || |||||| ||||||||||||||||||| ||
Db    901 GGTGGATGCAGAGAGAGAGATGGAGAGGCACAGAGTATCAGGATGATGAAAGAGTTTGG       960

Qy    961 AATAGGTGACCTCGTCGTGTGGGGAAAGATCAAGGGCTTCTCCTGGTGGGCCTGCCCATGGTGGT  1020
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AATAGGTGACCTCGTGTGGGGAAAGATCAAGGGCTTCTCCTGGTGGGCCTGCCCATGGTGGT     1020

Qy   1021 GTCCTGGAAAGCCACCTCCAAGCGACAGGCCATGCCCGGAAATGCCGCCTGGGTTACAGTGGTT  1080
          || |||||||||||||||||||||||||||||||||| ||||||||||||| |||| |||||||
Db   1021 GTCCTGGAAAGCCACCTCCAAGCGACAGGCCATGCCCGGAAATGCCGCCTGGGTTACAGTGGTT  1080

Qy   1081 TGGTGCATGGCAAGTTTTCTGCTACCTTGACAATCTCTGCTGACAAAACTGGTGCTCTGGGCTGT  1140
          ||||||||||||||||||||||||||| |||||||| ||||||||||||||||||| |||||||
Db   1081 TGGTGATGGCAAGTTTTCTGAGATCTCTGACAAAACTGGTGCTCTGGGGGCTGTTCAG       1140

Qy   1141 CCAGCACTTTAATTCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTACCA      1200
          ||||||||| |||| |||||||||||||||||||||||| ||||||||||||||||||||||
Db   1141 CCAGCACTTTAATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTACCA      1200

Qy   1201 CACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGAACCTTCTCCAGCAGTC              1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGAACCTTCTCCAGCAGTC              1260

Qy   1261 ACTGGAGGACCAGCTGAAGCCCATGCTGGAGTGGGCCCACGGTGGCCGCAAGTGACCTACTGG   1320
          | |||||||||||||||||||||| ||| || ||| ||||||||||| |||||||||||||||
Db   1261 ACTGGAGGACCAGCTGAAGCCCATGCTGGAGTGGGCCCACGGTGGCCGCAAGCCTACTGG       1320

Qy   1321 GATCGAGGGGGCCTCAAAACCCAACAAGAGAGCAACCAGTCGGTTTAATAAGTCGAAGGTGCGTCG 1380
          ||||||||||||||||| ||||||||||| ||||||||||||||
Db   1321 GATCGAGGGGGCCTCAAAACCCAACAAGAGAGCAACCA-------------               1354

Qy   1381 TTCAGACAGTAGGAACTTAGAACCCAGGAACAAAAGTCGAAGACGGCACAAC              1440
          ||||||||||||||||||||
Db   1355 -----------------------GAGAACAAAAGTCGAAGACGGCACAAC              1380

Qy   1441 CAATGACTCTGCTGCTTCTGAGTCCCCCCACCCCAAGCGCCTCAAGACAAAATAGCTATGG   1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAATGACTCTGCTGCTTCTGAGTCCCCCCACCCCAAGCGCCTCAAGACAAATAGCTATGG    1440

Qy   1501 CGGGAAGGAACCGAGGGGAGGGAGGGAGGAGATGAGGAGGCCGAGAACGGACTTCTGAAAGTCACCAA 1560
          ||||||||| ||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db   1441 CGGGAAGGACCGAGGGGAGGGAGGGGAGGATGAGGAGGCCGAGAACGGACTTCTGAAAGTCACCAA   1500
```

```
Qy   1561 CAACAAGGGCAAATCTGGAAGACCCGGCTGTTTGTCCTGTGGAAAGAGAACCCTGTGTCCTT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CAACAAGGGCAAATCTGGAAGACCCGGCTGTTTGTCCTGTGGAAAGAGAACCCTGTGTCCTT 1560

Qy   1621 CCACCCCCCTCTTTGAGGGTGGGCCTCTGCTCAGAGTTGCCCGGGGATCGCTTCCTAGAGCTCTT 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CCACCCCCCTCTTTGAGGGTGGGCCTCTGCTCAGAGTTGCCCGGGGATCGCTTCCTAGAGCTCTT 1620

Qy   1681 CTACATGTATGATGAGGACGGCTATCAGTCCTACTGCACCGTTGCCTGTGTGAGGGCCGTGA 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTACATGTATGATGAGGACGGCTATCAGTCCTACTGCACCGTTGCCTGTGTGAGGGCCGTGA 1680

Qy   1741 ACTGCTTGCTGTTGCAGTAAACACAAGCTTGCCGACGATCTTCTTGTGTGGAGTGTCTGGAGGT 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ACTGCTTGCTGTGTGCAGTAAACACAAGCTGCTCCAGATGCTTCTTGTGTGGAGTGTCTGGAGGT 1740

Qy   1801 GCTGGTTGGGCCGCCAGGCCACAGCTGAGGATGGCCAAGACCCTGGAGCTGCTATAT 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCTGGTTGGGCCGCCAGGCCACAGCTGAGGATGGCCAAGACCCTGGAGCTGCTATAT 1800

Qy   1861 GTGGCCTCCCCTCAGCCGCTGCCATGGGGTGCTCCCGACGCAGAAGATTGGAACAATGCCGCCT 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GTGCCCTCCCCTCAGCCGCTGCCATGGGGTTCCTCCGACGCAGAAAGATTGGAACAATGCGCGCCT 1860

Qy   1921 GCAAGAGACTTCTTCACTACTACTGATCCTGACCTGGAAGAATTTTGAGCCACCCAAGTTGTACCC 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCAAGAGACTTCTTCACTACTACTGATCCTGACCTGGAAGAATTTTGAGCCACCCAAGTTGTACCC 1920

Qy   1981 AGCAAATTCCTGCGAGCCAAAAAGGAGGGCCCATTAGTAGCTCCCTGTCTCTTGTTTGCATGGAAATTGC 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AGCAAATTCCTGCAGCCAAAAAGGAGGGCCCATTAGAGTCCCTGTCTCTTCTCTTGTTTTGATGGAAATTGC 1980

Qy   2041 AACGGGGTACTTGGTGCTCTCAAGGAGTTGGGGTTATTAAAAGTGGAAAAAGTACATTGCCTCCCGA 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTTATTAAAAGTGGAAAAAGTACATTGCCTCCGA 2040

Qy   2101 AGTCTGTGTGCAGAGTCCATCGGCTGTGTGGGAACTGGTTAAAGCATGAGAGGCCAGATCAAAATATGT 2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AGTCTGTGTGCAGAGTCCATCGGCTGTGTGGGAACTGGTTAAAGCATGAGAGGCCAGATCAAAATATGT 2100

Qy   2161 CAATGACGTCCGGAAAATCACCAAGAAAAATATTGAAGAGTGGGGGCCCGTTGCAGTTGGT 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CAATGACGTCCGGAAAATCACCAAGAAAAATATTGAAGAGTGGGGGCCCGTTGCAGTTGGT 2160

Qy   2221 GATTGGTGGAAGCCCATGCAATGATCTCTCTAACGTCAAATCCTGCCCGCAAAAGGTTTATA 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 GATTGGTGGAAGCCCATGCAATGATCTCTCTAACGTCAAATCCTGCCCGCAAAAGGTTTATA 2220

Qy   2281 TGAGGGCACAGGAGAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCGCCCCCAA 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TGAGGGCACAGGAGAGGCTCTTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCGCCCCCAA 2280

Qy   2341 GGAGGGCGGACAACCGTCCCAATTCTTCTTCTGGGATGTTCGAGAAATGTTCGTGGCCATGAAAAGTGAA 2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GGAGGGCGGACAACCGTCCCAATTCTTCTTCTGGGATGTTCGAGAAATGTTTGTGGCCATGAAAAGTGAA 2340

Qy   2401 TGACAAGAAAGACATCTCAAGATTCCTCGGCATGTAACCCAGTGATGATGGGTAGGCCATCAA 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TGACAAGAAAGACATCTCAAGATTCCTGGCATGTAACCCAGTGATCGATCGATGCCATCAA 2400

Qy   2461 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCTACCCGGAATGAACAGGCC 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCTACCCGGAATGAACAGGCC 2460

Qy   2521 CGTGATGGCTTCAAAGAATGATAAGCTCGAGCTGCCTCGGAGTTCAGTAGGAC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCTACCCGGAATGAACAGGGGC 2455

Db   2456 ------ 2455

Qy   2581 AGCAAAGTTAAAGAAAGTGCAGACAATAACCAAGTCGAACTCCATCAGACAGGGCAA 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2456 ------ 2455

Qy   2641 AAACCAGCTTTTCCCTGTAGTCATGAATGGCAAGGACGACGTTTTGTGGTGCACTGAGCT 2700
Db   2456 ----- 2455
```

```
Db   2456 ----------------------------------------- 2455
Qy   2701 CGAAAGGATCTTCGGCGTTCGGCCTTCCCTGCTCACTACACGGACGTGTCCAAACATGGAGCCGCGGCGC 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2456 ---AGGATCTTCGGCGTTCGGCCTTCCCTGCTCACTACACGGACGTGTCCAAACATGGGCCGCGGCGC 2511

Qy   2761 CCGTCAGAAAGCTGCTGGGCAGGTCCTGGAGTGTACGGGTCATCAGACGACAACACCTGTTTGCCCC 2820
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2512 CCGTCAGAAAGCTGCTGGGCAGGTCCTGGAGTGTACGGGTCATCAGACGACAACACCTGTTTGCCCC 2571

Qy   2821 CTTGAAGGAGCTACTTTGCCTCTGTGAATAGTTCTACCCAGGACTGGGGAGGAGCTCTGGTAGGT 2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2572 CTTGAAGGAGCTACTTTGCCTGTGAATAGTTCTACCCAGGACTGGGGAGGAGCTCTGGGTCAGA 2631

Qy   2881 GCCAGTGCCCAGAGTCACCCCTCCCTGAAGGACACCTCACCTGTCCCCTTTTTAGCTCACC 2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2632 GCCAGTGCCCAGAGTCACCCCTCCCTGAAGGACACCTCACCTGTCCCCTTTTTAGCTCACC 2691

Qy   2941 TGTGTGGGGGCCCTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGAGCAAGACCTC 3000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2692 TGTGTGGGGGCCTCACATCACTGTACCTCAGCTTTCTCCTCAGTGGGAGCAGACAGACCTC 2751

Qy   3001 CTGGCCCCTTGCAGGGAGGCCCGGGTGCTCCCCCTCCGTGTGGCACAGCCTCAGACCTGGCCTGCT 3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2752 CTGGCCCCTTGCAGGGGGAGCCCGGGTGCTCCCCCTCCGTGTGGCACAGCCTCAGACCTGGCCTGCT 2811

Qy   3061 TAGAGTAGCCCGGCATGGTGCTGCTGCATGTTCTCTCTTACCCTGAAACTTTAAAACTTGAAGTAG 3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2812 TACAGTAGCCCGGCATGGTGCTCATGTTCTTCTTTACCCTGAAACTTTAAAACTTGAAGTAG 2871

Qy   3121 GTAGTAAGATAGGCTTTCTTTCTTTTGTGGAAATGTGGGGCTTCAGATGAGCAAAGGCGTAAGAT 3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2872 GTAGTAAGATAGGCTTTCTTTTCTTTTGTATACTTCCTGAGTTTTATCACTCAGAAGTGATGGGCTAAGAT 2931

Qy   3181 ACCAAAAAAACAAAACAGAAAACCTCAACAGCTCTCTTA 3240
          |||||||||||||||||||||||||||||||||||||||
Db   2932 ACCAAAAAAACAAAACAGAAAACCTCAACAGCTCTCTTA 2991

Qy   3241 GTACTCAGGTTCATGCTGCAAAATCACTTGAGATTTTTGTTTTTTAAGTAACCGTGCTCCA 3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2992 GTACTCAGGTTCATGCTGCAAAATCACTTGAGATTTTTGTTTTTTAAGTAACCGTGCTCCA 3051

Qy   3301 AAAAATTCCCCCTCTCCCCCAGGAGTATTTGAAGATGATGATTTTATGGTTTAAGTCTTCC 3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3052 AAAAATTCCCCCTCTCCCCCAGGAGTATTTGAAGATGATGATTTTATGGTTTAAGTCTTCC 3111

Qy   3361 AAAAATTCCCCCTCTCCCCCAGGAGTATTTGAAGATGATGTTTAAGTCTTCC 3420
          ||||||||||||||||||||||||||||||||||||||||||||
Db   3112 AAAAATTCCCCCTCTCCCCCAGGAGTATTTTGAAGAGATGTTTTTATGGTTTTAAGTCTTCC 3171

Qy   3421 TGGCACCTTCCCCTTGCTTTGGTACAAGGGGTGAAGTCCCTGTTGGTCTTGTAGCAATTCC 3480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3172 TGGCACCTTCCCCTTGCTTTGGTACAAGGGGTGAAGTCCCTGTTGGTCTTGTAGCATTTCC 3231

Qy   3481 CAGGATGATGATGTCAGCAGGGATGACATCACCACCTTTAGGGGCTTTTCCCTGGCAGGGG 3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3232 CAGGATGATGATGTCAGCAGGGATGACATCACCACCTTTAGGGGCTTTTCCCTGGCAGGGG 3291

Qy   3541 CCCATGTGGCTAGTCCTCACGAAAGACTGGAGTAGAATGTTGGGAGCTCAGGAAGGGTGGG 3600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3292 CCCATGTGGCTAGTCCTCACGAAAGACTGGAGTAGAATGTTGGGAGCTCAGGAAGGGTGGG 3351

Qy   3601 TGGAGTGGGCCCTCTTCCAGGTGGTGGAGGGATACCGAAGGAGGAGAAAATCCATTC 3660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3352 TGGAGTGGGCCCTCTTCCAGGTGTTGAGGGATACGAAGGAGAAAATCCATTC 3411

Qy   3661 CCCACTCCCTCTTGCCAAAATGAGGGGCCCAGTCCCCAAACAGCTCAGGTGCCCCAGAACCCC 3720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3412 CCCACTCCCTCTTTGCCCAAAATGAGGGGCCCAGTCCCCAAACAGCTCAGGTGCCCCAGAACCCC 3471

Qy   3721 CTAGTTCCTCATGAGAAGCTAGGACCCAGAAGGACACATCGTTCCCCTTATCTCCGAGCAGTGTT 3780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 10

```
LOCUS       AF151971               4089 bp    mRNA    linear   ROD 08-JUN-2000
DEFINITION  Mus musculus DNA cytosine-specific methyltransferase isoform 3
            (Dnmt3b) mRNA, complete cds.
ACCESSION   AF151971
VERSION     AF151971.1  GI:8347122
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4089)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE     Cloning of full-length Dnmt3b cDNA and its alternative splicing
            isoforms in mouse embryonic tissue
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4089)
  AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
            Shen,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
            Biology, Institution of Basic Medical Sciences, Chinese Academy of
            Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
            P.R.China
FEATURES             Location/Qualifiers
     source          1..4089
                     /organism="Mus musculus"
                     /strain="KM"
                     /db_xref="taxon:10090"
                     /dev_stage="8-9 day old embryo"
     gene            1..4089
                     /gene="Dnmt3b"
     CDS             412..2742
                     /gene="Dnmt3b"
                     /function="methylates cytosine in DNA"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA cytosine-specific methyltransferase isoform
                     3"
                     /protein_id="AAF74517.1"
                     /db_xref="GI:8347123"
                     /translation="MKDSRHLNEEBGASGYEECIIVNGNFSDQSSDTKDAPSPVLE
                     AICTEPVCTEPTRGRRSSSSRLSKREVSSLLNYTQDMTGDGDRDDEVDDGNGSDILMPK
                     LTRETKDTRTRSESPAVRTRHSNGTSSLERQRASPRITRGRGRHHVQEYPVEPPATR
                     SRRRASSSSASTPWSSPASVDMEEVTPKSVSTPSVDLSQDGQDEGMDTTQVDAESIY
                     GDSTEYQDDKEFGIGDLVWGKIKGFSWMPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                     SEISADKLVALGLFSQHFNLATFNKLVSSRKAMVHTLERARVRAGKTFSSSPGESLED
                     QLKPMLEWAHGGFKPFGIEGLKPNKKQPENKSRRTTNDSASASEPPPKRLKTNSYGG
                     KDRGEDEESRERMASEYTNNKGNLBDRCLSCGKNPVSFHPLFEGGLCQSCRDFLEL
                     FYMIDDGIYQSYCTVCCEGRELLLCSNTSCCRCPCVECLDLVLVGACTAEDARLOEPWS
                     CYMCLPQRCHGVLRRRKDWNMRLLQDFFTTDPDLEEFEPPKLYPAIPAAKRRPIRVLSL
                     FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYNDVRKITKKNIEE
                     WGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYHLLNYTRPKEGDNRPFFWMF
                     ENVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARYFWGNLPGMNRIFGEPAHYTD
                     VSNMQRGARQKLLGRWSVSPVIRHLFAPLKDYFACE"
BASE COUNT      995 a   1071 c   1104 g    919 t
ORIGIN
```

Query Match 87.4%; Score 3664.8; DB 10; Length 4089;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 3935; Conservative 0; Mismatches 2; Indels 249; Gaps 2;

```
Qy     10 GCGCCCGGGGTTAAGCGGCGCCAAGTAAACGTAGGCGACGGATCGGCGGCGGGATTCGCGA 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GCGCGCGGGGTTAAGCGGCGCCAAGTAAAACGTAGGCGACGGATCGGCGGCGGGA 212

Qy     70 ACCCGACACTCCGGCGCGCGCCGCGCCGCCGGGCGGCCGCCGGGCGGCGCGCGGCGGCCGCGGGCGGCCGCTA 129
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 ACCCGACACTCCCGCGCGCCCGGCGCCAGGACCCCGGGCGGCGCGCGGCGGCGGCGGCTA 272

Qy    130 CAGCCAGCCTCACGACAGAGGCCCGGTGAGGCGTTTGTGCCAGACCTTGGAAAACCTCAGGTATA 189
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 CAGCCAGCCTCACGACAGAGGCCCGCTGAGGCTTGTGCCAGACCTTGGAAACCTCAGCAC 332

Qy    190 TACCTTTCCAGACGCGGGATCTCCCCTCCCCCATCCATAGTGCCCTTGGGACCAAATCCAG 249
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 TACCTTTCCAGACGCGGGATCTCCCCTCCCCCATCCATAGTGCCCTTGGGACCAAATCCAG 392

Qy    250 GGCCTTCTTTCAGGAAAACAANATGAAAGGGAGGAGACAGCAGACATCTGAAATGAACATCTGAAATGAAGAAGAGGGTGC 309
          |||||||||||||||||||||||||||||||||||||||||
Db    393 GGCCTTCTTTCAGGAAAACAANTGAAAGGGAGGAGACAGCAGACATCTGAAATGAAGAAGAGGGTGC 452

Qy    310 CAGCGGGGTATGAGGAGTGCATTATGTCGTTAATGTGGAACTTCAGTGACCAGTCCTCAGACAC 369
          |||||||||||||||||||||||||||||||||||||||||||
Db    453 CAGCGGGTATGAGGAGTGCATTATCGTTAATGGGAACTTCAGTGACCAGTCCTCAGACAC 512

Qy    370 GAAGGGATGCTCCCTCACCCCCACCAGTCTTGACGGCGACAATCTGCACAGACAGAGCCAGTCTGCACAC 429
          |||||||||||||||||||||||||||||||||||||||
Db    513 GAAGGGATGCTCCCTCACCCCCAGTGTTGGAGGCAAATCTGACAGAGCCAGTCTGCACAC 572

Qy    430 AGAGACCAGAAGGGCCGCAGGTCAAGCTCCCGGCGTCGTCTAAGAGGAGGAGTCCAGCCTTCT 489
          |||||||||||||||||||||||||||||||||||||||||
Db    573 AGAGACCAGAGGGCCGCAGGTCAAGCTCCCGGCTGTCTAAGAGGAGGAGTCTCCAGCCTTCT 632

Qy    490 GAATTACACGCAGGACATCACAGAGATGAGACAGATGANGAAGTAGAGATCAGATGGGAA 549
          |||||||||||||||||||||||||||||||||||||||
Db    573 GAATTACACGCAGGACATGACAGAGATGAGACAGAGATGAAGTAGAGATCAGATGGGAA 692

Qy    550 TGGCTCTGGATATTCTAAATGCCAAAAGCTCACCCGTGAGACCAAGGACACCAGGAAGGCAAAG 609
          |||||||||||||||||||||||||||||||||||||||
Db    693 TGGCTCTGGATATTCTAAATGCCAAAAGCTCACCCGTGAGACCAAGGACACCAGGAAGGCTC 752

Qy    610 TGAAAAGCCCCGGCTGTGTCCGAAACCACACATACCACGCTTGGGAGAGGGCAAAG 669
          |||||||||||||||||||||||||||||||||||||||
Db    753 TGAAAAGCCCCGGCTGTGTCCGAAACCCGACATAGCAATGGGACCTTGGGAGAGGCAAAG 812

Qy    670 AGCCTCCCCAGAATCACCCGAGTCGGCAGGGCCGCCCACCATGTCGCAAAGTGCAGGAGTACCCGTGT 729
          |||||||||||||||||||||||||||||||||||||||
Db    813 AGCCTCCCCAGAATCACCCGAGTCGGCAGGGCCGCCCACCATGTCGCAAGGAGTACCCCGTGT 872

Qy    730 GGAGTTTCCGGCGTACCAGTCTCGAGCATCGTCGAGCACGTCTTCACGCAAGCACGCCATG 789
          |||||||||||||||||||||||||||||||||||||||
Db    873 GGAGTTTCCGGCGTACCAGTCTCGAGCATCGTCTTCACGCAAAGCACGCCATG 932
```

```
Qy    790 GTCATCCCTGCCAGCCTCGACTTCATGGAAGAAGTGACACCTAAGAGCGGTCAGTACCCC  849
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 GTCATCCCTGCCAGCGGTCGACTTCAATGGAAGAAGTGACACCTAAGAGCGGTCAGTACCCC  992

Qy    850 ATCAGTTGACTTGAGCCAGGATGGAGAGATCAGGAGGGGTATGGATACCACACAGGTGGATGC  909
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 ATCAGTTGACTTGAGCCAGGATGGAGAGATCAGGAGGGGTATGGATACCACACAGGTGGATGC  1052

Qy    910 AGAGAGCAGAGATGGAGACAGCACAGAGAGTAATCAGGATGATGATAAAGAGTTTGGAATAGGTGA  969
          ||||||| | |  | |||||
Db   1053 AGAGAGCATATATGGAGACAGCACACAGAGTAATCAGGATGATCAGGAGTTTGGAATAGGTGA  1112

Qy    970 CCTCGTGTGGGGAAAGATCAAGGGGCTTCTCCTGGTTGGCCCTGCCCATGGTGGTGGTGTCCTGGAA  1029
          ||||||||||||||||||||||||
Db   1113 CCTCGTGTGGGGAAAAGATCAAGGGCTTCTCCTGGTTGGCCCCATGGTGGTGGTGTCCTGGAA  1172

Qy   1030 AGCCACCTCCAAGCGACACAGGCCCATGCCCGGGAAATGCGGCGTGGGTACAGTGGTTTGGTGATGG  1089
          ||||||||||||||||||||||||||||
Db   1173 AGCCACCTCCAAGCGACACAGGCCCATGCCCGGGAAATGCGGTTACAGTGGTTTGGTGATGG  1232

Qy   1090 CAAGTTTTCTGAGATCTCTGCTGACAAAACTGGTGCGGCTCTGGGGCGTTCAGCCAGCACTT  1149
          ||||||||| |||||||||||| |||||||
Db   1233 CAAGTTTTCTGAGATCTCTGCTGACAAAACTGGTGCGGCTCTGGGGCGTTCAGCCAGCACTT  1292

Qy   1150 TAATCTGGCCTACCTTCAATAAGCTGGTTTCTTTATAGGAAGGCCCATGTACCACACTCTGCTGA  1209
          ||||||||||||||||||||||||||||||
Db   1293 TAATCTGGCCTACCTTCAATAAGCTGGTTTCTTTATAGGAAGGCCCATGTACCACACTCTGCTGA  1352

Qy   1210 GAAAGCCAGGGGTTCGAGCTGGCAAGACCTTCTCCACCAGTCCTCTGGAGAGTCACTGGGGAGGA  1269
          ||||||||||||||||||||||||||||||||
Db   1353 GAAAGCCAGGGGTTCGAGCTGGCAAGACCTTCTCCACCAGTCCTCTGGAGAGTCACTGGGGAGGA  1412

Qy   1270 CCAGCTGAAGCCCCATGCTGTGGAGTGGGCCCCACGGGTGGCCTTCAGAGCCTACTGGGGATCGAGGG  1329
          |||||||||||||||||||||||||||||||
Db   1413 CCAGCTGAAGCCCCATGCTGTGGAGTGGGCCCCACGGGTGGCCTTCAGAGCCTACTGGGGATCGAGGG  1472

Qy   1330 CCTCAAAACCCAACAAAGAAGCAACACCAGTGGTTAATAAGTCGGAAGGGTCGGTCGTCGTTCAGACAG  1389
          |||||||||||||||||||||||||||
Db   1473 CCTCAAAACCCAACAAAGAAGCAACACCA                                      1497

Qy   1390 TAGGAAACTTAGAACCCAGGAGACGGCACGGAGAACAAAAAATGTCGAGAGACGGCACAACCAATGACTC  1449
          |||||||||||||||||
Db   1498 ------------------GAGAACAAAAAGTCGAAGAACCGGACAACACCAATGACTC  1532

Qy   1450 TGCTGCTTCTGAGTCCCCCCACCCAAACCGCCTCAAGACAAATAAGCTATGGCGGGAAGGA  1509
          ||||||||||||||||||||||||||||||
Db   1533 TGCTGCTTCTGAGTCCCCCCACCCAAGCGCCTCAAGACAAATAGCTATGGCGGGAAGGA  1592

Qy   1510 CCGAGGGGAGGATGAGGAGAGCCGAGAACGGAGATGGGCTTCTGAAGTCACCAACAACAAGGG  1569
          |||||||||||||||||||||||||||||||
Db   1593 CCGAGGGGAGGATGAGGAGAGCCGAGAACGGAGATGGGCTTCTGAAGTCACCAACAACAAGGG  1652

Qy   1570 CAATCTGGAAGACCGCTGTTTGTCCTGTGTGGAAAGAAGAAACCCTGTGCTTGCCTTCCACCCCCT  1629
          |||||||||||||||||||||||||||||
Db   1653 CAATCTGGAAGACCGCTGTTTGTCCTGTGGAAAGAAGAAACCCTGTGCTTGCCTTCCACCCCCT  1712

Qy   1630 CTTTTGAGGGTGGGGCTCTCGTCAGAGTTGCCCGGGGATCGGTCTTCCTAGAGCCTCTTCTACACATGTA  1689
          ||||||||||||||||||||||||||||||
Db   1713 CTTTGAGGGTGGGGCTCTCTGTCAGAGTTGCCCGGGGATCGGTCTTCTCTAGAGCCTCTTCTACATGTA  1772

Qy   1690 TGATGAGGAGGGCTATCAGTCACTCACTTCACGCACGCGCTGTCTGCGTGTGAGGGCGCGTGAACTGCTGCT  1749
          ||||||||||||||||||||||||||||
Db   1773 TGATGAGGACGGGCTATCAGTCACTCACTTCACGCACGCGCTGTCTGCGTGTGAGGGCGCGTGAACTGCTGCT  1832

Qy   1750 GTGCAGTAACACAAGCCTGCTGCAGATGCTTCTGTGTGTGGAGGTGTCTGGAGGTGGGG  1809
          ||||||||||||||||||||||||
Db   1833 GTGCAGTAACACAAGCCTGCTGCAGATGCTTCTGTGTGTGGAGGTGTCTGGAGGTGGGG  1892

Qy   1810 CGCAGGCCACAGCCTGAGGATGCCAAGCGCTGCAGCAACCCTGCGAGGAACCCTGGAGCTGCTATATGTGCCTCCC  1869
          |||||||||||||||||||||
Db   1893 CGCAGGCACAGCTGAGGATGCCAAGCTCCAAGGAACCCTGGAGCTGCGCTCGTATAATCTGCCTCCC  1952
```

```
Qy   1870 TCAGCCGCTGCCATGGGGGTCCTCGACGGCAGGAGGAGAAGATTGGAACATGCCGCCCTGCCAAGACTT  1929
          |||||||||||||||||||||||
Db   1953 TCAGCGCCTGCCATGGGGGTCCTCCGACGGCAGGAGAAAGATTGGAACATGCCGCCCTGCCAAGACTT  2012

Qy   1930 CTTCACTACTCAATCCTGACGTGGGAAGAAATTTGACGCAGGAAAATTGGAACTGGACCAGCAATTCC  1989
          ||||||||||||||||||||||||||
Db   2013 CTTCACTACTCAGATCCCTGACGTGGGAAGAAATTTGAGCGCACCCCAAGTTGTACCAGCAATTCC  2072

Qy   1990 TGCAGCCAAAAGGAGGCCCCATTAGACTCCGTCTCTGTCTTTGATGGAAGGTGCAACGGGTA  2049
          |||||||||||||||||||||||
Db   2073 TGCAGCCAAAAGGAGGCCCCATTAGAGCTCCGTCTCTTGTTTTGATGGAAGGTGCAACGGGTA  2132

Qy   2050 CTTGGGTGCTCAAGGAGTGTGGGTATTAAAAGTGGGAAAAAGTACAATTGCCTCCGAAGTCTGTGC  2109
          |||||||||||||||||||||||||
Db   2073 CTTGGGTGCTCAAGGAGTGTGGGTATTAAAAGTGGGAAAAAGTACAATTGCCTCCGAAGTCTGTGC  2132

Qy   2110 AGAGTCCATCGGCTGTGGGAACTGTTAAGCATGCATGAAGGCCAGATCAAATAATGTCAATGACGT  2169
          ||||||||||||||||||||||||||
Db   2133 AGAGTCCATCGGCTGTGGGAACTGTTAAGCATGCATGAAGGCCAGATCAAATAATGTCAATGACGT  2192

Qy   2170 CCGGAAAATCACCAAGAAAAAATATTGAAGAGTGGGGGCCCCGGTTTGTAATATGTCAATGACGT  2229
          |||||||||||||||||||||||||||
Db   2193 AGAGTCCATCGGCTGTGGGGAACTGTTAAGCATGAAGGCCAGATCAAATAATGTCAATGACGT  2252

Qy   2230 AAGCCCATGCAAATGAATGTCTCTCTAACGTCACAATCCTGCCCCGAAAAGGTTTATATGAGGGCAC  2289
          ||||||||||||||||||||||||||
Db   2253 CCGGAAAATCACCAAGAAAAAATATTGAAGAGTGGGGGCCCCGGTTCGACTTGGTGTGGTGG  2312

Qy   2290 AGGAAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCCGCCCCAAGAGGGGGCGA  2349
          ||||||||||||||||||||||||||
Db   2313 AAGCCCATGCAAATGAATGTCTCTCTAACGTCACAATCCTGCCCCGAAAAGGTTTATATGAGGGCAC  2372

Qy   2350 CAACCGTCCATTCTTCTTCTGGATGTTCTGGAGAATGTTGTGGCCATGAAAGTGAAATGACACAGAA  2409
          ||||||||||||||||||||||||||
Db   2373 AGGAAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCCGCCCCAAGAGGGGGCGA  2432

Qy   2410 AGACATCTCAAGATTCCTGGCATTGAATCTGGATGTTCGGAGAATGTTGTGCCATGAAAGTGAATGACAAGAA  2469
          ||||||||||||||||||||||||||
Db   2433 CAACCGTCCATTCTTCTTCTGGATGTTCTGGAGAATGTTGTGGCCATGAAAGTGAAATGACACAGAA  2492

Qy   2470 TGCTCACAGGGCGCCCCGGGTACTTCTTCTGGGGTAACCTACCCCGGAATGAAC  2529
          |||||||||||||||||||||||||||
Db   2493 AGACATCTCAAGATTCCTGGCATTGAATCTGGATGTTCGGAGAATGTTGTGCCATGAAAGTGAATGACAAGAA  2552

Qy   2530 TTCAAAGAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTCAGTAGGACAGCAAAGTT  2589
          ||||||||||||||||||||||
Db   2553 TGCTCACAGGGCGCCCCGGGTACTTCTTCTGGGGTAACCTACCCGGAATGAAC----------  2598

Qy   2590 AAAGAAAGTGCAGACAATAACCACCAAGTCGAAACTCCATCAGACACGGGCAAAAACCAGCT  2649
          |||||||||||||||||||||||
Db   2599 TTCAAAGAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTCAGTAGGACAGCAAAGTT  2598

Qy   2650 TTTCCCTGTAGTCATGAATGGCAGAAGGACGAGACGTTTTGTGGGTCCACTGGAGCTCGGAAAGGAT  2709
          |||||
Db   2599 -------------------------------------------------------------AGGAT  2603

Qy   2710 CTTCGGCCTTCCCTGCTCACTACACGGACGACTGTCCCAACCATGGGGCCGCGGGGGCCCGTCAGAA  2769
          ||||||||||||||||||||||||||
Db   2604 CTTCGGCGTTCCCTGCTCACTACACGGACGACTGTCCCAACATGGGGCCGCGGGGGCCCGTCAGAA  2663

Qy   2770 GCTGCTGGGCGCAGGTCCTGGAGTGTACCGGTCATCAGACACACCGTGTTTGCCCCCTTGAAGGA  2829
          ||||||||||||||||||||||||
Db   2604 CTTCGGCGTTCCCTGCTCACTACACGGACGACTGTCCCAACATGGGGCCGCGGGGGCCCGTCAGAA  2663

Qy   2830 CTACTTTGCCTGTGTGAAATAGTTCTACCCAGGACTGGGGAGGCTCGGGTCAGAGCCAGTGCC  2889
          ||||||||||||||||||||||||
Db   2724 CTACTTTGCCTGTGTGAAATAGTTCTACCCAGGACTGGGGAGGCTCGGGTCAGAGCCAGTGCC  2783

Qy   2890 CAGAGTCACCCCTCCCTGAAGGCACCTCACCTGTCCCCTTTTTAGCTCACCTGTGTGGGG  2949
          ||||||||||||||||||||||||
Db   2784 CAGAGTCACCCCTCCCTGAAGGCACCTCACCTGTCCCCTTTTTAGCTCACCTGTGTGGGG  2843

Qy   2950 CCTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGAGGAGAGAGCCTCCTGGCCCTT  3009
          ||||||||||||||||||||||||
Db   2950 CCTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGAGGAGAGAGCCTCCTGGCCCTT  3009
```

```
Db  2844  CCTCACATCACTGTACCTCAGCTTCTTCCTGCTCAGTGGGAGCAGAGAGCCTCCTGGCCCTT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3010  GCAGGGGAGCCCCGGTGCTGCTCCCCGGTGCACACCTGCTTGCTTAGAGTAGC  3069
Qy  2904  GCAGGGGAGCCCCGGTGCTGCTCCCCGGTGCACACCTGCTTGCTTAGAGTAGC  2963
Qy  3070  CCGGCATGGTGGCTCATGGTCTCATGGTTCTCTTCACCTGAAACTTTAAAACTTGAAGTAGGTAGTAGGA  3129
Db  2964  CCGGCATGGTGTGCTCATGTCTCATGTTCTCTTCACCTGAAACTTTAAAACTTGAAGTAGGTAGTAAGA  3023
Qy  3130  TGGGCTTTCTTTTTACCCTCCTGAGTTTATCACTCAGAGAAGTGGCTAAGATACCAAAAAA  3189
Db  3024  TGGGCTTTTCTTTTTACCCTCCTGAGTTTATCACTCAGAGAAGTGATGGCTAAGATACCAAAAAA  3083
Qy  3190  ACAAACAAAAAACAGAAACAGAAACAAAAAAACCTCAACAGCTCTCTTAGTTACTCAGG  3249
Db  3084  ACAAACAAAAAACAGAAACAGAAACAAAAAAACCTCAACAGCTCTCTTAGTTACTCAGG  3143
Qy  3250  TTCATGCTGCAAAATCACTGAGATTTGTTTTGTTTTTTAAGTAACCGTGCTGCTCCCACATTTGCTG  3309
Db  3144  TTCATGCTGCAAAATCACTGAGATTTGTTTTGTTTTTTAAGTAACCGTGCTGCTCCACATTTGCTG  3203
Qy  3310  GAGGAGTGCTATTGTGAATGTGGGCCTCAGAGTGACGATGAGCAAAGGGGGCAAAAAAAAAAATTCC  3369
Db  3204  GAGGAGTGCTATTGTGAATGTGGGCCTCAGAGTGAGCAAGGTCAAAGGGGCCAAAAAAAATTCC  3263
Qy  3370  CCCTCTCCCCCCCAGGAGTATTGAAGATGAGATGAGTTGAAGATGAGTTTAAGTCTTCCTGGCACCTT  3429
Db  3264  CCCTCTCCCCCCCAGGAGTATTGAAGATGAGATGAGTTTATGGTTTAAGTCTCTTCCTGGCACCTT  3323
Qy  3430  CCCCTTGCTTTGGTACAAGGGGCTGAAGTCCTTGTGTCTGTTGGTCTCTTGTAGCATTTCCCAGGATGAT  3489
Db  3324  CCCCTTGCTTTGGTACAAGGGGCTGAAGTCCTTGTGTCTGTTGTAGCATTTCCCAGGATGAT  3383
Qy  3490  GATGTCAGCAGGGATGACATCACCACCTTTAGGGCTTTTCCCTGGCAGGGGCCCATGTGTGG  3549
Db  3384  GATGTCAGCAGGGATGACATCACCACCTTTAGGGCTTTTCCCTGGCAGGGGCCCATGTGTGG  3443
Qy  3550  CTAGTCCTCACCAAAGACTGGAGTAGAAATGTTTTGGAGCTCAGAGAAGGGTGGGGGAGTGGC  3609
Db  3444  CTAGTCCTCACCGAAGACTGCTGAGTAGAAATGTTTGGAGCTCAGAGAAGGGTGGGGGAGTGGC  3503
Qy  3610  CCTCTTCCAGGTGTGAGGGGATACGAAGGAGGAGAAGCTTAGGGAAAAATCCATTCCCCACTCCC  3669
Db  3504  CCTCTTCCAGGTGTGAGGGGATACGAAGGAGGAGAAGCTTAGGGAAAAATCCATTCCCCACTCCC  3563
Qy  3670  TCTTTGCCAAATGAGGGGCCCAGTCCCCACAGTCCCCAGAGAACCCCCCAGTTCCT  3729
Db  3564  TCTTGCCAAATGAGGGGCCCCAGTCCCCACAGTCCCCAGAGAACCCCCTAGTTCCT  3623
Qy  3730  CATGAGAAGCTTAGGACCAGAAGCACACATCGTTCCCCTTATTCTGAGCAGTGTTTGGGGAACT  3789
Db  3624  CATGAGAAGCTTAGGACCAGAAGCACACATCGTTCCCCTTATCTGAGCAGTGTTTGGGGAACT  3683
Qy  3790  ACAGTGAAAAACCTTCTGGAGATGTTAAAAGCTTTTTACCCCACGATAGAATTGTGTTTTTA  3849
Db  3684  ACAGTGAAAAACCTTCTGGAGATGTTAAAAGCTTTTTACCCCACGATAGAATTGTGTTTTTA  3743
Qy  3850  AGGGGTGCTTTTTTTAGGGGCATCACTGGAGATAAGAAAGTGCATTTCAGAAAATGCCAT  3909
Db  3744  AGGGGTGCTTTTTTTTAGGGGCATCACTGGAGATAAGAAAGTGCATTTCAGAAAATGCCAT  3803
Qy  3910  CGTAAGTGGTTTTTTTAAACACCTTTTACCTAATTACAGGTGCTGCTATTTTATAGAACCAGACAA  3969
Db  3804  CGTAAGTGGTTTTTTTAAACACCTTTTACCTAATTACAGGTGCTGCTATTTTATAGAACCAGACAA  3863
Qy  3970  CACTTCTTTTATTGACTCTCAGACTTCATTCTATTTTTGTAACTCGC  4029
Db  3864  CACTTCTTTTTATTGACTCTCAGACTTCATGTTACCATTTTTTTTTGTAACTCGC  3923
Qy  4030  AAGGTGTGGGCTTTTGTAACTTCAACAGGTGTGGGGAGAGACTGCCTTGTTCTTGTTTCAACAGTTT  4089
Db  3924  AAGGTGTGGGCTTTGTAACTTCAACAGGTGTGGGGAGAGACTGCCTTGTTCTTGTTTCAACAGTTT  3983
Qy  4090  GTCTCCACTGGTTTCTTTCTAATTTTTTAGGTGCAAAAGATGACAGATGCCCAGAGTTTACCTTTC  4149
Db  3984  GTCTCCACTGGTTTCTTCTTAATTTTTTAGGTGCAAAAGATGACAGATGCCCAGAGTTTACCTTTC  4043
Qy  4150  TGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAA  4195
Db  4044  TGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAA  4089
```

RESULT 11

FEATURES   Location/Qualifiers
     source   1..4034
                     /organism="Mus musculus"
                     /strain="KM"
                     /db_xref="taxon:10090"
                     /dev_stage="8-9 day old embryo"
     gene   1..4034
                     /gene="Dnmt3b"
     CDS   297..2687
                     /gene="Dnmt3b"
                     /function="methylates cytosine in DNA"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA cytosine-specific methyltransferase isoform
                     7"
                     /protein_id="AAF74521.1"
                     /db_xref="GI:8347135"
                     /translation="MKGDSRHLNEEGASGYEECIIVNGNFSDQSSDTKDAPSPPVLE
                     AICEPVCTETRGRNSSSSRLSKREVSSLLNYTQDMTGDGDRDDEVDDGGNGSDILMPK
                     LTREFKDTRTRSESPAVRTRHSNGTSSLERQRASPRITRGRQGRHIHVQEYPVEPFATR
                     SRRRASSSSASTPWSSASVDEMEEVTPRKSVSTPSVDLSQDGDQEGMDTTQVDAESIY
                     GDSTEYQDDKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                     SEISADKLVALGLFSQHFNLATFNKLVSYRKAMHTFLEKARVRAGKTFSSSPGESLED
                     QLKPMLEWAHGGFKPTGIEGLKPNKKQPVVNKSKVRRSDSRNLEPRRERNKSRRRTTN
                     DSAASRSPPPRKLKTNSYGKRGRGEDBEESRERBMASEVTNNKGNLEDDRCLSGCKKNPVS
                     FHPLEGGLCQSCRDRFELEFYMDEDGYQSYCTVCCEGRELLLCSNTSSCCRCFVEC
                     LEVLVGAGTAEDAKLQBPWSCYMCLPQRCHGVLRRKDWNMRLQDFFTTDDDLEEFEP
                     PKLIPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKH
                     EGQIKYVNDVRKITKKNIEEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTCRLFFEFY
                     HLLNYTRPKEGDNRPFFWMFENVVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARY
                     FWGNLPGMNRIFGFPANYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"

BASE COUNT   991 a   1042 c   1097 g   904 t
ORIGIN

Matches 3880; Conservative 0; Mismatches 2; Indels 304; Gaps 2;

```
Qy     10 GCGCCCGGGGGTTAAGCGCGCCCCAAGTAAACGTAGCGCCAGCCAGCCCGGAGATTCGCGA 69
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 GCGCCCGGGGGTTAAGCGCGCCCCAAGTAAACGTAGCGCCAGCCAGCCCGGAGATTCGCGA 69

Qy    153 GCGGCCCGGGGGTTAAGCGCGGCCCCAAGTAAACGTAGCGCAGCGATTCGCGA 212
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GCGGCCCGGGGGTTAAGCGCGGCCCCAAGTAAACGTAGCGCAGCGATTCGCGA 212

Qy     70 ACCCGACACTCCGCGCGCCGCCCGCCCGCCCGCCCAGACCCGCGGCGGCCGCGGCGCTA 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 ACCCGACAACTCCCGCGCGCCGCCCGGCGCCGAGCGCCCCCGGATCGGCGGCCCCGCGCTA 272

Qy    130 CACCCAGCCTCACGACACAGGCCCGGCTGAGGCGTTGTGTCGCAGACCTTGGAAACCTCAGGTATA 189
          ||||||||||||||||||||||||||||
Db    273 CAGCCAGCCTCACGA----------- 287

Qy    190 TACCTTTCCAGACGCGGGGATCTCCCTCCCCCATCCATAGTTGCCTTGGGACCAAAATCCAG 249
          ----------------------------
Db    288 ----------------------------- 287

Qy    250 GGCCTTCTTTCAGGAAAACAAATCAAAGGGAGACAGCACACATCTGAATGAGAAGAAGAGGGTGC 309
          -------------
Db    288 -----CAGGAAAACAAATGAAAGGGAGACAGCACAGAATCTGAATGAATGAAGAAGAGGGTGC 337
```

(sequence alignment continues)

```
Qy   2170 CCGGAAAATCACCAAGAAAAAATATTGAAGAGAGTGGGGCCGTTCGGTTCGACTTGGTGATTGGTGG 2229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2198 CCGGAAAATCACCAAGAAAAAATATTGAAGAGAGTGGGGCCGTTCGACTTGGTGATTGGTGG 2257

Qy   2230 AAGCCCATGCAATGATCTCTCTAACGTCAATTCCTGCCCCGGAAAGGTTTATATGAGGGCAC 2289
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2258 AAGCCCATGCAAATGATCTCTCTAACGTCAATCCTGCCCCGGAAAGGTTTATATGAGGGCAC 2317

Qy   2290 AGGAAGGGCTCTTCTTCGGAGTTTTTTACCACTTGCTGAAGTTATACCCGGCCCCAAGGAGGGGGA 2349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2318 AGGAAGGGCTCTTCTTCGAGTTTTTACCACTTGCTGAAATTATACCCGGCCCCAAGGAGGGGGGA 2377

Qy   2350 CAACCGTCCAATTCTTCTTCGGAGTTTCGGAGAAATGTTGTGGGCCATGAAATGTTCAACAGAA 2409
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2378 CAACCGTCCAATTCTTCTTCGGAGATTTCGGAGAAATGTTGTGGGCCATGAAATGGACAAGAA 2437

Qy   2410 AGACATCTCAAGATTCCTGGCATTGTAACCCAGTGATGAATCGAATTCAAGGTGTCTGC 2469
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2438 AGACATCTCAAGATTCCTGGCATTGTAACCCAGTGATGAATCAAGGTGTCTGC 2497

Qy   2470 TGCTCACAGGGCCCGTGTACTTCTTGGGGTAAACCCGAAATGACACAGGGCCCGTGATGGC 2529
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2498 TGCTCACAGGGCCCGTGTACTTCTGGGGTAAACCTACCCGGGAATGAAC 2543

Qy   2530 TTCAAAGAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTCAGTAGGACAGCAAAGTT 2589
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2543 --------------------------------------------- 2543

Qy   2544 ------------------------------------------------- 2543
Db   2544 --------------------------------------------- 2543

Qy   2590 AAAGAAAAGTGCAGACAAATAACCACCAGTCGAACTCCATCAGACAGGGCAAAAACCAGCT 2649
Db   2544 --------------------------------------------- 2543

Qy   2650 TTTCCCTGTAGTCATGAAATGGCAAGGACGACGTTTTGTGGTGCACTGAGCTCGAAAGGAT 2709
          |||||
Db   2544 -----AGGAT 2548

Qy   2710 CTTCGGGCTTCCCTGCTCACTCACGCAGTGTCCAACATGGGCCCGGGCCCGTCAGAA 2769
          ||||||
Db   2549 CTTCGGCTTCCCTGCTCACTACACGCAGTGTCCAACATGGCCGGGCCCGTCAGAA 2608

Qy   2770 GCTGCTGGGGCAGGGTCCTGGAGGTCATCAGAGACAACCTGTTTGCCCCCTTGAAGGA 2829
          ||||||
Db   2609 GCTGCTGGGGCAGGGTCCTGGAGGTGTACCGGGTCATCAGAGACAACCTGTTTGCCCCCTTGAAGGA 2668

Qy   2830 CTACTTTGCCCGTGAAGAGTAGTTCACCAGGAGCTCTCGGTCACGAACTCAGGTCCC 2889
          ||||||
Db   2669 CTACTTTGCCTGTGAAGAGTAGTTCACCAGGAGCTCTCGGTCACGAACTCAGGTCCC 2728

Qy   2890 CAGAGTCACCCCTCCCTGGAAGGGCACCTCACCTGTCCCCTTTTTTAGCTCACCTGTGTGTGGGG 2949
          ||||||
Db   2729 CAGAGTCACCCCTCCCTGAAAGGCACCTCACCTGTCCCCTTTTTTAGCTCACCTGTGTGGGG 2788

Qy   2950 CCTCACATCACTGTACCTCAGCTTTCTTCTCCTGCTCAGTGGGAGCAGAGCCTCCTGGCCCTT 3009
          ||||||
Db   2789 CCTCACATCACTGTACCTCAGCTTTTCTCCTGCTCAGTGGGAGCAGAGCCTCCTGGCCCTT 2848

Qy   3010 GCAGGGGAGCCCGGTTGCTCCCTCCGTGGTGCACAGCTCAGACCTGCCTGCTGTTAGAGTAGGC 3069
          ||||||
Db   2849 GCAGGGGAGCCCGGTTGCTCCCTCCGTGTGCACAGCTCAGACCTGCCTGCTGTTAGAGTAGGC 2908

Qy   3070 CCGGCATGGGTGCTCATGTTCTCTTTACCCTGAAACTTTTAAAACTTGAAGTAGGTAGTTAGA 3129
          ||||||
Db   2909 CCGGGCATGGGTGCTCATGTTCTCTTTACCCTGAAACTTTTAAAACTTGAAGTAGGTAGTTAGA 2968

Qy   3130 TGGGCTTTCTTTTACCCTCCTGAGTTTATCACTCAGAAGTGATGGCTAAGGATACCAAAAA 3189
          ||||||
Db   2969 TGGCTTTTCTTTTACCCTCCTGAGTTTATCACTCAGAAGATGGCTAAGGATACCAAAAA 3028

Qy   3190 ACAAACAAAAACAGAAACAAAAAAACCTCAACAGACGTCTCTTAGTTAGCTACTCAGG 3249
          ||||||
Db   3029 ACAAACAAAAAACAGAAACAAAAAAACAAAAAAACCTCAACAGCCTCTTCTTAGTTAGCTACTCAGG 3088

Qy   3250 TTCATGCTGCAAAAATCACTTGAGATTTTGTTTTTAAGTAACCCGTGCTGCTCCACATTTGCTG 3309
Db   3089 ---------------------------------------------
```

```
Db   3089 TTCATGCTGCAAAAATCACTTGAGATTTTGTTTTTAAGTAACCCGTGCTGCTCCACATTTGCTG 3148

Qy   3310 GAGGAATGCTATTGTGAGATGTGGGCTCAGATGAGATGGAGCAAAGGTCAGAGGGCCAAAAAAATTCC 3369
          ||||||
Db   3149 GAGGAATGCTATTGTGAGATGTGGGCTCAGATGAGATGGAGCAAAGGTCAGAGGGCCAAAAAAATTCC 3208

Qy   3370 CCCTCTCCCCCCAGGAGTATTTGAAGATGATGATGGTTATTGTTTAGTCTTCCTGGCCACCTT 3429
          ||||||
Db   3209 CCCTCTCCCCCCAGGAGTATTTGAAGAAGATGATGATGGTTATTGTTTAGTCTTCCTGGCCACCTT 3268

Qy   3430 CCCCTTGCTTTGGGTACAAAGGGCTGAAGTGCCTGTTGGTCTTGGTGTAGGCATTTCCCAGGATGAT 3489
          ||||||
Db   3269 CCCCTTGCTTTGGGTACAAAGGGCTGAAGTCCTGTTGGTCTTGGTGTAGGCATTTCCCAGGATGAT 3328

Qy   3490 GATGTCAGCAGGGATGAGACATCACCACCTTTAGGGGCTTTTCCCTTGCCAGGGGCCCCATGTGG 3549
          ||||||
Db   3329 GATGTCAGCAGGGATGAGACATCACCACCTTTCCCTTGCCAGGGGCCCCATGTGG 3388

Qy   3550 CTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAGGTGGGGTGGGAGTGGC 3609
          ||||||
Db   3389 CTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAGGTGGGGTGGGAGTGGC 3448

Qy   3610 CCTCTTCCAGGTGTGAGGGATACGAAGGAGGATCAGGAGGGAAAATCCATTCCCCACTCCC 3669
          ||||||
Db   3449 CCTCTTCCAGTGTGAGGGATACGAAGGAGGAAACCTTAGGGAAATCCATTCCCCACTCCC 3508

Qy   3670 TCTTGCCAAATGAGAGGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGGAACCCCCTAGTTCCT 3729
          ||||||
Db   3509 TCTTGCCAAATGAGAGGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGGAACCCCCTAGTTCCT 3568

Qy   3730 CATGGAAAGCTAGGACCAGAAGCACACATCGTTCCCCTTATCTGAGCAGTGTTTGGGGAACT 3789
          ||||||
Db   3569 CATGAGAAAGCTAGGACCAGAAGCACACATGTTCCCCTTATCTGAGCAGTGTTTGGGGAACT 3628

Qy   3790 ACAGTCGAAAAACCTTCTCGGAGATGTGAAAGCTTTTTTACCCCACGATAGAGTTGTTTTTTTA 3849
          ||||||
Db   3629 ACAGTGAAAAACCTTCTCGGAGATGTTAAAAGCTTTTTTACCCCACGATAGAGTTGTTTTTTA 3688

Qy   3850 AGGGGGTGGCTTTTTTGTAACTCAGGACTTCCTAATTTCAGGTGTCATGTTTTTGTAACTGCGC 3909
          ||||||
Db   3689 AGGGGGTGGCTTTTTTGTAACTCAGACTTCCTAATTTCAGGTGTCATGTTTTTGTAACTGCGC 3748

Qy   3910 CGTAATGGTTTTTAAACAACCTTTTACCTAATTACAGGTGTCTATTTTTATAGAGAGCAGAA 3969
          ||||||
Db   3749 CGTAATGGTTTTTAAACAACCTTTTACCTAATTACAGGTGTCTATTTTTATAGAGAGCAGAA 3808

Qy   3970 CACTTCTTTATTGACTCTCAGACTTCTAATTTCCATTTTTTTGTTAACTCGC 4029
          ||||||
Db   3809 CACTTCTTTATTGACTCTCAGACTTCTAATTTCCATTTTTTTGTTAACTCGC 3868

Qy   4030 AAGGTGTGGGGCTTTTTGTAACTTCACAGGTGTGGGGGAGAGACTGCCTTGTTCAACAGTTT 4089
          ||||||
Db   3869 AAGGTGTGGGGCTTTTTGTAACTCACAGGTGTGGGGGAGAGACTGCCTTGTTCAACAGTTT 3928

Qy   4090 GTCTCCACTGGTTTCTAATTTTTAGGTGCAAAGATGACAGATGCCCAGGAGTTTACCTTTC 4149
          ||||||
Db   3929 GTCTCCACTGGTTTCTAATTTTTAGGTGCAAAGATGACAGATGCCCAGGAGTTTACCTTTC 3988

Qy   4150 TGGTTGATTAAAGTTGTATTTTCTCTAAAAAAAAAAAAAAAAAAA 4195
          |||||||||||||||||||
Db   3989 TGGTTGATTAAAGTTGTATTTTCTCTAAAAAAAAAAAAAAAAAAAAAAA 4034
```

```
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1  (bases 1 to 3974)
AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
          Shen,Y.
TITLE     Cloning of full-length Dnmt3b cDNA and its alternative splicing
          isoforms in mouse embryonic tissue
JOURNAL   Unpublished
REFERENCE  2  (bases 1 to 3974)
AUTHORS   Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
          Shen,Y.
TITLE     Direct Submission
JOURNAL   Submitted (17-MAY-1999) Department of Biochemistry and Molecular
          Biology, Institution of Basic Medical Sciences, Chinese Academy of
          Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
          P.R.China
FEATURES          Location/Qualifiers
     source       1..3974
                  /organism="Mus musculus"
                  /strain="KM"
                  /db_xref="taxon:10090"
                  /dev_stage="8-9 day old embryo"
     gene         1..3974
                  /gene="Dnmt3b"
     CDS          297..2627
                  /gene="Dnmt3b"
                  /function="methylates cytosine in DNA."
                  /note="alternatively spliced"
                  /codon_start=1
                  /product="DNA cytosine-specific methyltransferase isoform
                  8"
                  /protein_id="AAF74522.1"
                  /db_xref="GI:8347137"
                  /translation="MKGDSRHLNEEEGASYEECIIVNGNFSDQSSDTKDAPSPVLE
                  AICTEPVCTEPTRGRRSSSRLSKREVSSLLNYTQDMTGDGDRDEVDDGNGSDILMPK
                  LTREYKDTPTRPTRSESPAVRTRHSNGTSSSLERQRASPRITRGRGGRHHVQEYPVEFPATR
                  SRRRASSSASSTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDGDQEHGMDTTQVDAESIY
                  GDSTEYQDDKEFGIGDLVWGKIKGFSWWPAMVVSWKAVTSKRQAMPGHMRVQWFGDGKF
                  SEISADKLVALGLFSQHFNRLATFRNKLVSYRKAMYHTLEKARVRAGKTFSSSPGESLED
                  QLKPMLEWANGGFKPTGIEGLKPNKKQPENKSRRTTNDSAASESPPPKRLKTNSYGG
                  KDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVSFHPLFEGGLCQSCRDRFEL
                  FYMYDEDGYOSYCTVCCEGRELLLCSNTSCCRPCVECLEVILGVAGTAEDAKLQEPWS
                  CYMCLPQRCHGVLRRRKDWNMRLODFFTDPDLEEFEPPKLYPAIPAAKRRIRVLSL
                  FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYVNDVRKITKKNIEE
                  WGPFDLVIGGSPCNDLSNVNPARRGLYEGTGRLFFEFYHLLNYTRPKEGDNRPFFWMF
                  ENVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARYFWGNLPGMNRIFGFPAHYTD
                  VSNMGRGARQKLLGRSWSVPVIRILFAPLKDYFACE"
BASE COUNT    972 a    1032 c    1078 g    892 t
ORIGIN
```

```
1st Local Similarity     81.6%;     Score 3424.8;     DB 10;     Length 3974;
Matches 3689; Conservative    93.5%; Pred. No. 0;   Mismatches  7;  Indels 249;  Gaps  2;
          0; Mismatches
```

```
Qy       251 GCCTTCTTTCAGGAAAACAAATGAAGGGAGACAGCAGCAGACATCTGAATGAAGAAGAGGCTGCC 310
             |||||
Db       519 AATTACACGCAGGACATGACAGGAGATGGAGACACAGAGGATGAAGTAGATGGGGAAT 578

Qy       279 GCCTCACGACAGGAAAACAAATGAAGGGAGACAGCAGCAGACATCTGAATGAAGAAGAGGTGCC 338
             |||||
Db       551 GGCTCTGATATTCTAATGCCAAAGCTCACCGTGACAAACCAAGGAGAACCACAGAGACGACGTCT 610

Qy       311 AGCGGGGTATGAGGAGTCCATTATCGTTAATGGGAACTTCAGTGACCAGTCCTCAGACACG 370
             |||||
Db       579 GGCTCTGATATTCTAATGCCAAAGCTCACCGTGAGAACCAAGGAGAACCACAGAGACGACGTCT 638

Qy       339 AGCGGGTATGAGGAGTCCATTATCGTTAATGGGAACTTCAGTGACCAGTCCTCAGACACG 398
             |||||
Db       611 GAAAGCCCGGCTGTCCGAACCCGACCAGAACCCGAACATAGCAATGGGAACTCCAGCTTGGAGGGCAAAAGA 670

Qy       371 AAGGATGCTCCCTCACCCCACCCCAGTCTTGGAGGCAAATCTGCACAGAGCCCAGTCTGCACACCA 430
             |||||
Db       639 GAAAGCCCGGCTGTCCGAACCCGACCAGAACCCGAACATAGCAATGGGAACTCCAGCTTGGAGGGCAAAAGA 698

Qy       399 AAGGATGCTCCCTCACCCCAGTCTTGGAGGCAAATCTGCACAGAGCCCAGTCTGCACACCA 458
             |||||
Db       671 GCCTCCCCCAGAATCACCCGAGGTCGGGCCAGGGCCGGCGGGCCCACCCATGTCGCAGGAGTACCCTGTG 730

Qy       491 AATTACACGCAGGACATGACAGGAGATGGAGACACAGAGGATGAAGTAGATGGGGAAT 550
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1032 GCCTCCCCCAGAATCACCCGAGGTCGGGCCAGGGCCGGCGGGCCCACCCATGTCGCAGGAGTACCCTGTG 758

Qy       431 GAGACCAGAGGCCGCAGGTCAAGCTCCCGGCGGTTGTGTCTAAGAGGGAGGTCTCCAGCCTTCTG 490
             |||||
Db       699 GCCTCCCCCAGAATCACCCGAGGTCGGGCCAGGGCCGGCGGGCCCACCCATGTCGCAGGAGTACCCTGTG 790

Qy       459 GAGACCAGAGGCCGCAGTCAAGCTCCCGGCTGTCTCAAGAGGGAGGTCTCCAGCCTTCTG 518
             |||||
Db       731 GAGTTTCCGGCTACCAGGTCTCGGGAGACGTCGGGGCCCGGAGGCCGGCGTGCTTCAGCCAAGCAACACCACCATGG 818

Qy                                                                                 519
Db       759 GAGTTTCCGGCTACCAGGTCTCCGGAGACGTCGAGCATCGTCGTTCAGCAAAGCAACACCACCATGG 818
```

```
Dt       791 TCATCCCCCTGCCAGCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCA 850
Qy       519 AATTACACGCAGGACATGACAGGAGATGGAGACACAGAGGATGAAGTAGATGGGGAAT 578
Qy       551 GGCTCTGATATTCTAATGCCAAAGCTCACCGTGACAAACCAAGGAGAACCACAGAGACGACGTCT 610
Dt       819 TCATCCCCCTGCCAGCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCA 878
Qy       579 GGCTCTGATATTCTAATGCCAAAGCTCACCGTGAGAACCAAGGAGAACCACAGAGACGACGTCT 638
Db       851 TCAGTTGACTTGAGCCAGGATGGAGATCAGGAGCAGGGTATGGATACCACACAGGTGGATGCA 910
Qy       611 GAAAGCCCGGCTGTCCGAACCCGACCAGAACCATAGCAATGGGAACTCCAGCTTGGAGGGCAAAAGA 670
Db       879 TCAGTTGACTTGAGCCAGGATGGAGATCAGGAGCAGGGTATGGATACCACACAGGTGGATGCA 938
Qy       639 GAAAGCCCGGCTGTCCGAACCCGACCAGAACCATAGCAATGGGAACTCCAGCTTGGAGGGCAAAAGA 698
Db       911 GAGAGCAGAGATGGAGAGCACAGAGTAGTCAGGATGATGAATAAAGAGAGTTTGGAATAGGTGAC 970
Qy       671 GCCTCCCCCAGAATCACCCGAGGTCGGGCCAGGGCCGGCGGGCCCACCCATGTCGCAGGAGTACCCTGTG 730
Db       939 GAGAGCATATATGGACAGCACAGAGTATCAGGATCATAAAGAGTTTGGAATAGGTGAC 998
Qy       699 GCCTCCCCCAGAATCACCCGAGGTCGGGCCAGGGCCGGCGGGCCCACCCATGTCGCAGGAGTACCCTGTG 758
Db       971 CTCGTGTGTGGGGAAAGATCAAAGGGCTTCTCCTGGTGGGCCTGCCATGGTGGTTCCTGGAAA 1030
Qy       731 GAGTTTCCGGCTACCAGGTCTCGGGAGACGTCGGGGCCCGGAGGCCGGCGTGCTTCAGCCAAGCAACACCACCATGG 790
Db       999 CTCGTGTGTGGGGAAAGATCAAAGGGCTTCTCCTGGTGGGCCTGCCATGGTGGTTCCTGGAAA 1058
Qy       759 GAGTTTCCGGCTACCAGGTCTCCGGAGACGTCGAGCATCGTCGTTCAGCAAAGCAACACCACCATGG 818
Db      1031 GCCACCTCCAAAGCGACAGGCCATGCCCGGAGAATGCGTCGGGGTACAGTGGTTTGGGTGATGGC 1090
Qy       791 TCATCCCCCTGCCAGCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCA 850
Db      1059 GCCACCTCCAAAGCGACAGGCCATGCCCGGAGAATGCGTCGGGGTACAGTGGTTTGGGTGATGGC 1118
Qy       819 TCATCCCCCTGCCAGCGTCGACTTCATGGAAGAAGTGACACCTAAGAGCGTCAGTACCCA 878
Db      1091 AAGTTTTCTGAGAATCTCTGCTGACAAACTGGTTGGGCTCTGTGGGGCTTGTTCAGCCAGCACTTT 1150
Qy       851 TCAGTTGACTTGAGCCAGGATGGAGATCAGGAGCAGGGTATGGATACCACACAGGTGGATGCA 910
Db      1119 AAGTTTTCTGAGAATCTCTGCTGACAAACTGGTCGGCTCTGTGGGGCTTGTTCAGCCAGCACTTT 1178
Qy       879 TCAGTTGACTTGAGCCAGGATGGAGATCAGGAGCAGGGTATGGATACCACACAGGTGGATGCA 938
Db      1151 AATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAGAGGCCATGTACCACACTCTGGAG 1210
Qy       911 GAGAGCAGAGATGGAGAGCACAGAGTAGTCAGGATGATGAATAAAGAGAGTTTGGAATAGGTGAC 970
Db      1179 AATCTGGCTACCTTCAATAAGCTGGTTTTCTTATAGGAGAAGGCCATGTACCACACTCTGGAG 1238
Qy       939 GAGAGCATATATGGACAGCACAGAGTATCAGGATCATAAAGAGTTTGGAATAGGTGAC 998
Db      1211 AAAGCCAGGGTTCGAGCGTGGCAAGAGACCTTCTCCAGCAGAGTCCTGGAGAGTCACTGGAGGAC 1270
Qy       971 CTCGTGTGTGGGGAAAGATCAAAGGGCTTCTCCTGGTGGGCCTGCCATGGTGGTTCCTGGAAA 1030
Db      1239 AAAGCCAGGGTTCGAGCGTGGCAAGAGACCTTCTCCAGCAGAGTCCTGGAGAGTCACTGGAGGAC 1298
Qy       999 CTCGTGTGTGGGGAAAGATCAAAGGGCTTCTCCTGGTGGGCCTGCCATGGTGGTTCCTGGAAA 1058
Db      1271 CAGCTGAAGCCCCATGCCTGGAGTGGGCCCACGCGGTTAATAAGTCGAAGGTCGTTCAGACAGT 1330
Qy      1031 GCCACCTCCAAAGCGACAGGCCATGCCCGGAGAATGCGTCGGGGTACAGTGGTTTGGGTGATGGC 1090
Db      1299 CAGCTGAAGCCCATGCCTGGAGTGGGCCCACGCGGTTAATAAGTCGAAGGTCGTTCAGACAGT 1358
Qy      1059 GCCACCTCCAAAGCGACAGGCCATGCCCGGAGAATGCGTCGGGGTACAGTGGTTTGGGTGATGGC 1118
Db      1331 CTCAAAACCCAACAAGAAGCAACCAGTGGTTAATAAGTCGAAGGTCGTTCAGACAGT 1390
Qy      1091 AAGTTTTCTGAGAATCTCTGCTGACAAACTGGTTGGGCTCTGTGGGGCTTGTTCAGCCAGCACTTT 1150
Db      1359 CTCAAAACCCAACAAGAAGCAACC------------------------- 1381
Qy      1119 AAGTTTTCTGAGAATCTCTGCTGACAAACTGGTCGGCTCTGTGGGGCTTGTTCAGCCAGCACTTT 1178
Db      1391 AGGAAACTTAGAACCCAGGAGACGGCGGGAACAAAAAGTCGAAGACGACAACAACAATGACTCT 1450
Qy      1151 AATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAGAGGCCATGTACCACACTCTGGAG 1210
Db      1382 ---------------------------------------CGAGAACAAAAAGTCGAAGACGACAACAACAATGACTCT 1418
Qy      1179 AATCTGGCTACCTTCAATAAGCTGGTTTTCTTATAGGAGAAGGCCATGTACCACACTCTGGAG 1238
Db      1451 GCTGCTTCTGAGTCCCCCCCCCACCCAAGCGCCTCAAGACAAAATAGCTATTATGGGGGAAGGAC 1510
Qy      1211 AAAGCCAGGGTTCGAGCGTGGCAAGAGACCTTCTCCAGCAGAGTCCTGGAGAGTCACTGGAGGAC 1270
Db      1419 GCTGCTTCTGAGTCCCCCCCACCCAAGCGCCTCAAAGACAAAATAGCTATTATGGGGGAAGGAC 1478
Qy      1239 AAAGCCAGGGTTCGAGCGTGGCAAGAGACCTTCTCCAGCAGAGTCCTGGAGAGTCACTGGAGGAC 1298
Db      1511 CGAGGGGAGGATGAGGAGAGCCGAGAACGGATGGGCTTCTGAAGTCACCAACAACAGGGC 1570
Qy      1271 CAGCTGAAGCCCCATGCCTGGAGTGGGCCCACGCGGTTAATAAGTCGAAGGTCGTTCAGACAGT 1330
Db      1479 CGAGGGGAGGATGAGGAGAGCCGAGAACGGATGGGCTTCTGAAGTCACCAACAACAGGGC 1538
Qy      1299 CAGCTGAAGCCCATGCCTGGAGTGGGCCCACGCGGTTAATAAGTCGAAGGTCGTTCAGACAGT 1358
Db      1571 AATCTGGAAGAACCGCTCTGTTTGTCCTCTTGGGAAAGAAGAAAACCCTGTGTGTCCTCCACCCCCTC 1630
Qy      1331 CTCAAAACCCAACAAGAAGCAACCAGTGGTTAATAAGTCGAAGGTCGTTCAGACAGT 1390
Db      1539 AATCTGGAAGAACCGCTCTGTTTGTCCTCTTGGGAAAGAAGAAAACCCTGTGTGTCCTCCACCCCCTC 1598
```

```
Qy   1631 TTTGAGGGTGGGCTTCTGTCAGAGTTGCCGGGGATCGGCTCTGTCAGAGCTCTTCTACATGTAT 1690
          |||||||||||||||||||||||||||||||| ||||||  ||| |||||||||||||||||||
Db   1599 TTTGAGGGTGGGCTTCTGTCAGAGTTGCCGGGATCGGCTTCCTAGAGCTCTTCTACATGTAT 1658

Qy   1691 GATGAGGACGGCGCTATCAGTTCCTACTGCACCGTGTGCTGTGTGAGGGGCGTGGTGCTGCTG 1750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 GATGAGGACGGCGCTATCAGTTCCTACTGCACCGTGTGCTGTGTGAGGGGCGTGGTGCTGCTG 1718

Qy   1751 TGCAGTAACACAAGCTGCTGCAGAATGCTTCTGTGTGGAGTGTCGGAGGTGGTGGTGGGC 1810
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 TGCAGTAACACAAGCTGCTGCAGATGCTTCTGTGTGGGAGTGTCTGGAGTGCTGGGTGGGGC 1778

Qy   1811 GCAGGCACACGCTGAGAATGCCAAGCTGCAGGAAACCTGCAGGACGCTGCTATATGTGCCCCT 1870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 GCAGGCACACGCTGAGAATGCCAAGCTGCAGGAAACCTGCAGGACGCTGCTATATGTGCCCCT 1838

Qy   1871 CAGCGCCTGCCATGGGGGTCCTCCGACGCACGCAGGGAAAAGATTGGAACATGGGCACTGGC 1930
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 CAGCGCCTGCCATGGGGGTCCTCCGACGCACGCAGGGAAAAGATTGGAACATGGGCACTGGC 1898

Qy   1931 TTCACTACTGACCTGAAGAATTGAGCCACCCAAGTTGTACCCACCAAGTTCCT 1990
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1899 TTCACTACTGACCTGAAGAATTGAGCCACCCAAGTTGTACCCACCAAGTTCCT 1958

Qy   1991 GCAGGCCAAAAGGAGGCCCATTAGAGTCCTGGTGCTCTGTGTTGATGGAATTGCAACGGGGTAC 2050
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1959 GCAGGCCAAAAGGAGGCCCATTAGAGTCCTGGTGCTCTGTGTTGATGGAATTGCAACGGGGTAC 2018

Qy   2051 TTGGTGCTCAAGGAGTTGGGTATTAAAGTGGAAAAGTACAATTGCCTCCCGAAGTCTGTGCA 2110
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2019 TTGGTGCTCAAGGAGTTGGGTATTAAAGTGCCTCCCGAAGTCTGTGCA 2078

Qy   2111 GAGTGCCATCGGTGTGGGGAAACTGTTAAGCATGACAGCCAGATGCTCAATGACGTC 2170
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2079 GAGTGCCATCGGTGTGGGGAAACTGTTAAGCATGACAGCCAGATGCTCAATGACGTC 2138

Qy   2171 CGGAAAATCACCAAGAAAAATAATTGAAGAGTGGGGCGCCCGTTCGACTTGGTGATTGGTGGA 2230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2139 CGGAAAATCACCAAGAAAAATAATTGAAGAGTGGGGCGCCCGTTCGACTTGGTGATTGGTGGA 2198

Qy   2231 AGCCCATGCAATGATCTCTCTAACGTCAATCCTGCCCGCAAAGGTTTATAATGACGTC 2290
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2199 AGCCCATGCAATGATCTCTCTAACGTCAATCCTGCCCGCAAAGGTTTATAATGACGTC 2258

Qy   2291 GGAAGGCTCTTCTTCTTCGAGTTTTTACCACTTTGCTGAATTATACCCGCCCAAGGAGAGGGGAC 2350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2259 GGAAGGCTCTTCTTCTTCGAGTTTTTACCACTTTGCTGAATTATACCCGCCCAAGGAGAGGGGAC 2318

Qy   2351 AACCGTCCAATTCTTCTCTGGATTGTTCGGGCCATGATGAAATGATGACAAGAAA 2410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2319 AACCGTCCAATTCTTCTCTGGATTGTTCGGGCCATGATGAAATGATGACAAGAAA 2378

Qy   2411 GACACTCAAGATTCCTGGCATGTAAACCCAGTGAATGAAATCGATGCCAATCAAGGTTGTCTGCT 2470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2379 GACACTCAAGATTCCTGGCATGTAAACCCAGTGAATGAAATCGATGCCAATCAAGGTTGTCTGCT 2438

Qy   2471 GCTCACAGGGGCCCGGTACTTTCTGGGGTAACCTACCCGGGAATGAAACAGGCCCGTGATGGGCT 2530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2439 GCTCACAGGGCCCGGTACTTTCTGGGGTAACCTACCCGGGAATGAAC---------------- 2483

Qy   2531 TCAAAGAATGATAAGCTCGAGCTGCCTGGAGTTCAGTAGGACAGCAGAAGTTA 2590
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   2484 -----------------------------------------------

Qy   2591 AAGAAAGTGCAGACAAATAACCACCAAAGTCGAAACTCCATCAGACAGGGCAAAAAACCAGCTT 2650
          ||||
Db   2484 ----------------------------------------- 2483

Qy   2651 TTCCCTGTAGTCATGAATGGTCGCACTGAGCTCGAAAGGATC 2710
          |||||||||||
Db   2484 ----------------------------AGGATC 2489
```

```
Qy   2711 TTCCGGCTTCCCTGCTCACTACACACGGACGTGTCCAACATGGGCGCGGGGGCCCCCTTGAAGGAC 2770
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2490 TTCCGGCTTCCCTGCTCACTACACACGGACGTGTCCAACATGGGCGCGGGGGCCCCCTTCAGAAG 2549

Qy   2771 CTGCTGGGCACGGTCCTGGAGTGTACCGGTGGTAGAGTGGGTGTACCGGTGTCAGTGTTGAAGGAC 2830
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2550 CTGCTGGGCACGGTCCTGGAGTGTACCGGTGGTAGAGTGGGTGTACCGGTGTCAGTGTTGAAGGAC 2609

Qy   2831 TACTTTGCCGTGTGAAATAGTTCTTACCCAGGACTGGGGAGTCTCGGTCAGACGCCC 2890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2610 TACTTTGCCGTGTGAAATAGTTCTTACCCAGGACTGGGGAGTCTCGGTCAGACGCCC 2669

Qy   2891 AGAGTGCACCCCTCCCTGAAAGGCACCTCACCTGTCCCCTTTTTAGCTGTACGCTGGGGC 2950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2670 AGAGTGCACCCCTCCCTGAAAGGCACCTCACCTGTCCCCTTTTTAGCTGTACCACCTGGGGC 2729

Qy   2951 CTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGGAGCCTCAGTGGGGAGCCTCCCTG 3010
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2730 CTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGGAGCCTCAGTGGGGAGCCTCCCTTG 2789

Qy   3011 CAGGGGAGCCCGGTGCTCCCCGTGTGCACAACTCAGAGCTGCTGCTTTAGAGTGTACC 3070
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2790 CAGGGGAGCCCGGTGCTCCCCGTGTGCACAACTCAGAGCTGCTGCTTTAGAGTGTACC 2849

Qy   3071 CGGGCATGGTGGCTCATGTTCTCTTTACCCTGAGTTTATCACTTCAGAAGTTGATGGGATCAGGAT 3130
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2850 CGGGCATGGTGGCTCATGTTCTCTTTACCCTGAGTTTATCACTTCAGAAGTTGATGGGATCAGGAT 2909

Qy   3131 GGCTTTCTTTACCCTCCTGAGTTTATCACTGAAGTTTGTTTTAAGTAACCCGTGCTCCACATTTGCTGG 3190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2910 GGCTTTCTTTACCCTCCTGAGTTTATCACTGAAGTTTGTTTTAAGTAACCCGTGCTCCACATTTGCTGG 2969

Qy   3191 CAAACAAAAACAGAAACAAAAAAACAAAACCTCAACAGCTCTCTTAGTACTCAGGT 3250
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2970 CAAACAAAAACAGAAACAAAAAAACAAAACCTCAACAGCTCTCTTAGTACTCAGGT 3029

Qy   3251 AGGAGGCTATTGTGAATGTGGGCTTCAGATGAGCAAGGTCAAGGGGCCAAAAAAATTCCC 3310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3030 AGGAGGCTATTGTGAATGTGGGCTTCAGATGAGCAAGGTCAAGGGGCCAAAAAAATTCCC 3089

Qy   3311 AGGATGCTATTGTGAATGTGGGCTTCAGATGAGCAAGGTCAAGGGGCCAAAAAAATTCCC 3370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3090 TCATGCTGCAAAAATCACTTGACATTTGTTTTAAGTAACCCGTGCTCCACATTTGCTGG 3149

Qy   3371 CCTCTCCCCCAGGAGTATTTGAAGATGATTTTGTTTTGTTTTAAGTTGTGG 3430
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3150 CCTCTCCCCCAGGAGTATTTGAAGATGATTTTGTTTTGTTTTAAGTTGTGG 3209

Qy   3431 CCCTTGCTTGGTACAAGGGGCTGAAGTCCTGTGTTGGTCTTGTAGCATTTCCCAGGATGATG 3490
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3210 CCCTTGCTTGGTACAAGGGGCTGAAGTCCTGTGTTGGTCTTGTAGCATTTCCCAGGATGATG 3269

Qy   3491 TAGTCCTCACGAAGACTGGAGTAGAAATGTTTGGGAGCTCAGGAGCCC 3550
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3270 ATGTCAGCAGGGATGACATCACCACACCTTTAGGGCTTTTCCCTGGCAGGGGCCATGTGGCC 3329

Qy   3551 TAGTCCTCACGAAGACTGGAGTAGAAATGTTTGGGAGCTCAGGAGCCC 3610
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3330 TAGTCCTCACGAAGACTGGAGTAGAAATGTTTGGGAGCTCAGGAGCCC 3389

Qy   3611 CTCTCTTCCAGGTGTGTGAGGGGATACGGAAGGGAGGGAAATCCATTCCCCACTCCCT 3670
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3390 CTCTCTTCCAGGTGTGTGAGGGGATACGGAAGGGAGGGAAATCCATTCCCCACTCCCT 3449

Qy   3671 CTTGGCCAAATGAGGGCCCAGTGCCCAACAGCTCAGGTCCCCAGAAATCAGGGTACTC 3730
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3450 CTTGGCCAAATGAGGGCCCAGTGCCCAACAGCTCAGGTCCCCAGAAATCCATTAGTTCCTC 3509

Qy   3731 ATGGAAAGCTAGGACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTTTGGGGACTA 3790
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3510 ATGGAAAGCTAGGACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTTTGGGGACTA 3569

Qy   3791 CAGTGAAAACCTTCTGGAGGATGTTAAAAAGCTTTTTTACCCCACGGATAGAATTGTGTTTTTAA 3850
```

RESULT 13
AF331857

LOCUS   AF331857   4335 bp   mRNA   linear   PRI 02-JAN-2002
DEFINITION Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA,
            complete cds.
ACCESSION   AF331857
VERSION   AF331857.1 GI:18033254
KEYWORDS   .
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4335)
  AUTHORS   Ni,J., Pradhan,S. and Roberts,R.J.
  TITLE   Cloning, expression and characterization of human DNMT3 genes
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 4335)
  AUTHORS   Ni,J., Pradhan,S. and Roberts,R.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
            Beverly, MA 01915, USA
FEATURES   Location/Qualifiers
     source   1..4335
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene   1..4335
                     /gene="DNMT3B"
     CDS   322..2883
                     /gene="DNMT3B"
                     /codon_start=1
                     /product="DNA cytosine methyltransferase 3 beta"
                     /protein_id="AAL57040.1"
                     /db_xref="GI:18033255"
                     /translation="MKGDTRHLNGEEDAGGREDSILVNGACSDQSDSDSPPILEAIRTP
                     EIRGRRSSSRLSKREVSSLLSYTQDLTGDGDGEDGDGSDTPVMPKLFRETRSRSESPA
                     VRTRNNSVSSRERHRPSPRSTRGRQGRNHVDESPVEFPATRSLRRRATASAGTPWPS
                     PPSSYLTIDLTDDTEDTHGTPQSSSTPYARLAQDSQGGMESPQVEADSGDDGSSEYQ
                     DGKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMSGMRWVQWFGDGKFSEVSADK
                     LVALGLFSQHFNLATEFNKLVSRKAMHIALEKARVRAGKTPSSSPGDDSLEDQLKPMLE
                     WAHGGFKPTGIEGLKPNNTQPVVNKSKVRRAGSRKLESRKYENKTRRRTADDSATSDY
                     CPAPKRLKTNCYNNGKDRGDEDQSREQMASDVANNKSSLEDGCLSCGRKNPVSFHPLF
                     EGGLCQTCRDRFLELFYMKDDGYQSYCTVCCEGRELLCGNTSCCRCFCVECLEVLV
                     GTGTAAEAKLQEPWSCYMCLPQRCHGVLRRRKDWNVRLQAFFTSDTGLEYEAPKLYPA
                     IPAARRRPIRVLSLFDGIATGYLVLKELGIKVGKYVASEVCEESIAVGTVKHEGNIKY
                     VNDVRNITKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEYHLLNYS

ORIGIN

Query Match   43.3%; Score 1815.4; DB 9; Length 4335;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 2534; Conservative 0; Mismatches 651; Indels 138; Gaps 19;

Qy   RPKEGDDRPFWMFENVVAMKVGDKRDISRFLEGNPVMIDAIKVSAAHRARYPWGNLP
Qy   GNRRPVIASKNDKLELQDCLEYNRIAKLLKRVQTIITTKNSIKQGKNQLFPVVWNGKED
Qy   VLWCTELERIPGFPVHYTDVSNMGRGAROKLLGRSWSVPVIRHLFAPLKDYFACE"

Db 3570 CAGTGAAAAACCTTCTGGAGAGTGATGTTAAAAGCTTTTTACCCCACCGATAGATTGTTGTTTTAA 3629
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3851 GGGGTGCTGCTTTTTTTAGGGGGCATCACTGGAGATAAGAGAAAGCTGCATTCAGAAATGCCATC 3910
Db 3630 GGGGTGCTGCTTTTTTTAGGGGCATCACTGGAGATAAGAAAAGCTGCATTCAGAAATGCCATC 3689
Qy 3911 GTAATGGTTTTTTAAACACCTTTTTTACCTAATTACAGGTGCTATTTTTATAGAAGCAGAACAAC 3970
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3690 GTAATGGTTTTTTAAACACCTTTTTTACCTAATTACAGGTGCTATTTTTATACAGAACAAC 3749
Qy 3971 ACTTCTTTTTTATGACACTCTCAGACACTTCTATTTTCATGTTTACCATTTTTTTTTGTAACTCGGCA 4030
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3750 ACTTCTTTTTTATGACACTCTCAGACTTCTATTTTCATGTTTACCATTTTTTTGTAACTCGCA 3809
Qy 4031 AGGTGTCGGGCTTTTGTAACTTCACACAGGTGTGGGGAGAGACTGCCTTGTTTGTTTCAACAGTTTG 4090
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3810 AGGTGTGGGGCTTTTGTAACTTCACACAGGTGTGGGGAGAGACTGCCTTGTTTGTTTCAACAGTTTG 3869
Qy 4091 TCTCCACTGGTTTCTAATTTTTTTAGGTGCAAAGATGCACAGATGCCCCAGAGAGTTTACCTTTTCT 4150
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3870 TCTCCACTGGTTTCTAATTTTTTAGGTGCAAAGATGACAGATGCCCAGAGAGTTTACCTTTCT 3929
Qy 4151 GGTTGATTAAAGTTGTATTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAAAAA 4195
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 3930 GGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3974

Qy   251 GCCTTCTTTTCAGGAAAACAAATGAAAGGGAGACAGCAGACAATCGAAATGAAGAAGAGGGTGCC 310
        |||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   304 GCCCTGGGCGCAGGAAAAGCATGAAAGGGAGACAACCAGACCATCTCAATGGAGAGGAGAGCGCC 363
Qy   311 AGCGGGTAIGAGAGGAGTGGCATTAIATCGTTAATCGTTAATGGGAGAACTTCAGTGACCAGTCCTCAGACACG 370
        ||||||||| ||| ||||||||||||| ||| ||||||||||||||||||| |||||
Db   364 GGCGGGAGGGAAAGACTCGGATACCTTCGTCAACGGGGCCGGTGCACCAGGACCAGTCCTCCGA---- 419
Qy   371 AAGGATGCTCCCTCACCCCCCAGTCTTGGAGGCAAATCTGCACAGAGCCAGTCTGCACACCA 430
        ||| |||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   420 -----------CTCGCCCCCAATCCTGGAGGCGTA------------TCCGCACCCCG 453
Qy   431 GAGAACCAGAGGCCGCGAGGTCAAGACTCCCCGGTCGTGTCTAAGAGGGAGGAGGTCTCCAGCCCTTCTG 490
        |||| ||||||||||||||||||| |||||||||||||| |||||||||||||||||||
Db   454 GAGATCAGAGGCCGAAGATCAAAGCTCGCGACTCTCCAAGAGGGAGGAGGTGTCCCAGTCGCTA 513
Qy   491 AATTACACGCAGGACACATGACACAGGAGATGGGAGAGACAGAGAGATGATGAAGTAGAATGAAGATGGGAAT 550
        |||||||||||||||||||||||||||||||||||| |||||||||||
Db   514 AGCTACACAGCAGGACACTTGACAGGCGATGGCGAC--------GGGGAAGAATGGGGAT 561
Qy   551 GGCTCTGATAATC---TAATGCCAAAAGCTCACCCGTTCGTGAGACCAAGGACACCGACGCGCC 607
        || |||| ||| |||||||||||| ||||||||||||||| |||||||||||
Db   562 GGCTCTGACACCCCAGTCATGCCAAAAGCTCTTCCG--------GGAAACCAAGGACTCGT 612
Qy   608 TCTGAAAAGCCCGGCTGTCCGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAA 667
        |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   613 TCAGAAAAGCCCACACGCTGTCGAACACTGGAAAATAAACAACAGTGTCTCCAAGCCCGAGAGGCAC 672
Qy   668 AGAGCCTCCCCAGAATCACCCGAAGGTCGGCAGGGCCGCGACCCACCATGTGCAGGAAGAGTACCCT 727
        ||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   673 AGGCCTTCCCCACGTTCCACCGAGGCCGCCGGAGGCGCCGCAAACCATGTGGACGAGTCCCC 732
Qy   728 GTGGAGTTTCCGGCTACCACACAGGTGCTCGGAGACGCTCGGGAGCCATCGTTCAGCAACGACAGCCA 787
        |||||| ||||||||||||||||||||||||||||||||||||| |||||||||||
Db   733 GTGGAGTTCCCGGCTACCAGGTCCCCTGAGACGGCGGCAACAGCAICGGCAGGGAAACGCCA 792
Qy   788 TGGTCATCCCCTGCCAGGTCGGACTTCATGGAAGAAGT--------------------- 825
        ||||||||||||||||||||||||||||||| |||
Db   793 TGGCCGGTCCCCTCCCAGCTCITTACCTTACCATCGACCTCACAGACGACAGGAGGACACA 852
Qy   826 ----GACACCTAAGAGCGTCAGTACCCCATCAGTTGACTTGAGCCGGATGGAGGATCAG 880
        ||||||| ||||||||||||||||| |||||||||||||||||
Db   853 CATGGGACGCGCCCCAGAGCAGTACCCCTCTAGCCGCCGGTGTGGGGCGGAAAGATCAAAGGGGCTTCTCC 912
Qy   881 GAGGGTATGGATTACCACACAGGTGGATGCAGAGGAGCGAGAGGATGGAGACAGCCACAGAGTAT 940
        |||| ||||||||||||||||||| |||||||||||||||||||||||||||
Db   913 GGGGGCATGAGGACTCCCCGCCGCAGGTGGACGACAGTGGCGACGGCAGCAGCGAGTTCAGAGTAT 972
Qy   941 CAGGATGATAAAAGAGTTTGGAAATAGGTGACCTCGTGTGTGGGGAAAGATCAAAGGGGCTTCTCC 1000
        ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   973 CAGGATGGGAAGGAGTTTGGAAATAGGGGACCTCGTGTGTGGGGCAAGATCAAAGGGTTCTCC 1032
Qy 1001 TGGTGGCCTGCCATGGTGGTCGTCCTGGAAAAGCCACCTCCAAGCGACAGCCACTGCCCGGA 1060
        ||||||||||||||||||| ||||||||||||| ||||||||||||||||||||||
Db 1033 TGGTGGCCCGCCATGGTGGTGTCTTCGGAAAGGCCACCTCCAAAGACGACAGCAGCGAGTAT 1092
Qy 1061 ATGCGCTGGGTACAGTGGTTTGGTGATGGGCAAGTTTTCTGAGATCTCCTGTGACAAACTG 1120
        ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db 1093 ATCCGGTGGGTCCAGTGGTTTGGGCGATGGCAAGTTCTCCGAGGTCTCTGCAGACAAAACTG 1152
Qy 1121 GTGGGCTCTGGGGCTGTTCAGCCAGCACTTTAATCTGGGCTACCTTCAAATAAGCTGGTTTCT 1180
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1153 GTGGCACTGGGGCTGTTCAGCCAGCACTTTAATTTGGCCACCTTCAATAAGCTCGTCTCC 1212

```
Qy    1181 TATAGGAAGGCCATGTACCACACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGACCTTC 1240
           ||| ||||||| || |||||||||  |||||||||||||||||||||||||||||||||||
Db    1213 TATCGAAAAGCCATGTACCATGCTCTGGAGAAAGCTAGGGTGCGAGCTGGCAAGACCTTC 1272

Qy    1241 TCCAGCAGTCCTGGGAGGTCACTGGAGTCACAGCTGAAGCCCATGCTCTGGAGTGGGCCCAC 1300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1273 CCCAGCAGCCCTGGGAGACTCATTGGAGGACCAGCTGAAGCCCATGTTGGAGTGGGCCCAC 1332

Qy    1301 GGTGGCTTCAAGCCTACTGGGATCGGAGGGCCTCAAACCCAACACAAGCAACCAGTGGTT 1360
           || |||||||||||| || |||||||||||||||||||||||||||||||||||||||||
Db    1333 GGGGGCTTCAAAGCCCCACTGGGATCGGAGGGCCTCAAACCCAACACGCAACCAGTGGTT 1392

Qy    1361 AATAAAGTCGAAGGTGCGTCGTGCCAGGCAGTAGGAAACTTAGAACCCGGAGAACGCGAGAAC 1420
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    1393 AATAAGTCGAAGGTCGTCGTGCCAGGCAGTAGAAATTTAGAATCAGAGGAAATACGAGAAC 1452

Qy    1421 AAAAGTCGAAGACGCAACCAAATGACTCTGCTTCTG---AGTCCCCCACCCAAG 1477
           |||  |||||||||||||||||||||||||||||||   || | ||||||||||
Db    1453 AAGACTCGAAAGACGCAACCAGCTGAACGACTCAGCCACCTCTGACTACTGCCC 1512

Qy    1478 CGGCCTCAAGACAAAATAGCTAT---GGGGGGAAGGACCCGAGGGGAGGGGGAGAGGAGCCCGA 1534
           ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1513 CGCCTCAAGACAAAATTGCTATAACAACGGCAAAGACCCGAGGGGATGAAGACATCAAGAGCCGA 1572

Qy    1535 GAACGGATGGCTTCTGAAGTCACCAACAACAAGGGCAATCTGGAAGACCCGCTGTTTGTCC 1594
           ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    1573 GAACAAATGGCTTCAGATGTTGCCAAACAAGCAGCCTGGGAGCTGGAAGAATGGCTCAAA 1632

Qy    1595 TGTGGAAAAGAAGAAGAACCCTGTGTGTCGTGTTGTGTCTGCAGTAACACAAGCTGCTGCGA 1654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1633 TGTGGCAGGAAAAAACCCGTGTCCTTCCACCCTCTTTGAGGGGGTGGGCTCTGTGTCAGACA 1692

Qy    1655 TGCCGGGATGCGCTTCCTAGAGCTCTTCTACATGTATGATGAGGACGGGCTATCAGTCCTAC 1714
           |||||||||||||| ||||||||||||||| || || ||||||||||||||||||||||||
Db    1693 TGCCGGGATCGGCTTCCTTGAGCTGTTTTACATGTATGCTGGCTATCAGTCTTAC 1752

Qy    1715 TGCACCGTGTTGCTGTCGTGTGAGGGCGTGAACTGCTGCTGGAGTAACACAAGCTGCTGCAGA 1774
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1753 TGTGACTGTGTGTGCTGGGAGGGCCGAGAGCTGCTGCGTGCCAGGCACGACGAGCTGCTGCCAG 1812

Qy    1775 TGCTTCTGTGTGGGAGTGTCTCGAGGTGCTGGTGGCGGCACACTGCCAAG 1834
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1813 TGTTTCTGTGTGGGAGTGTCCTGGAGGTGCTGGTGGGCACAGCGGCGAGGCCAAG 1872

Qy    1835 CTGGCAGGAACCCTGCGAGCGCTGCTAATATGTGCCTCCCTCCAGCGGCTGCCCATGGGGTCTCCGA 1894
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1873 CTTCAGGAGCGCCTGAGCGCTGTTACATGTGCTCCCGCACGCGTGTCATGCGGTCCTGCGG 1932

Qy    1895 CGCAGGAAAACAGATTGGAACATCGCGCCGCCTGCAAGACTTCTTCACTACTGCGACCTGGGAA 1954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1933 CGCCGGAAGGACTGGAACGTGCGCCCTGCAGGCCTTCTTCACCAGTGACACGTGACACGTGAA 1992

Qy    1955 GAATTGAGCCACCCAAGTTGTACCCACAGCAAATTCCTGCAGCCAAAAGGAGGCCCATTAGA 2014
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1993 TA---TGAAGCCCCAAAGCTGTACCCTGGTTTGCTGGCTCAGCCCGAAAGGAGGCGGCCATCGA 2049

Qy    2015 GTCCTGTCTCTGTTTGATGGAATTGCAACGGGGTTACTTGGTGCTCAAGGAGTTGGGTATT 2074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2050 GTCCTGTCATTGTTTGAATGGCATCGCGACACAGGTTACCTCCTCAAAAGATGGCTGTGGGTATA 2109

Qy    2075 AAAGTGGAAAAGTACATGTGCCTCCGAAGTCTGTGCGAGAGTCCATCCGCTGTGTGGGAACTGTT 2134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2110 AAGGTAGGAAAGTACGTCGGCTTCTGAAGTGTGTGAGGAGTCCATTGCTGTTGTTGGAACCGTG 2169

Qy    2135 AAGCATGAAGGCCCAGATCAAATATATGTCAAATGAGTCCGGAAAAAATCACCAGAAAAATATT 2194
           ||||| |||||||||||||||||||| |||| |||||||||||||||| |||||||||||||
Db    2170 AAGCACGAGGGGAATATCAAAAATACGTGAACAGACTCAAAGAAAAAATATT 2229

Qy    2195 GAAGAGTGGGGGCCGTTCGACTTGGTGATTGGGTGGAAGGCCCATGCAAATGAAATGATCTCTCTAAC 2254
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    2230 GAAGAATGGGGGCCCATTTGACTTGGTGATTGGCGGAAGCCCATGCAACGATCTCTCAAAT 2289
```

```
Db    2255 GTCAATCCTGCCCCGCAAAAGGTTTATATGAGAGGGGCCACAGGAAGGGCTCTTCTTCTTCGAGTTTTAC 2314
           |    |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2290 GTGAATCCAGCCAGGAAAAGGCCTTGTATGAGGGGTACACGGCCGGGCTTCTTCTTCGAATTTTAC 2349

Qy    2315 CACTTGCTGAAATTTATACCCCGCCCAAGGAGGGGCGCAGACAACCGTCCATTCTTCTCTGGAATGTTTC 2374
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    2350 CACCTGCTGAATTTACTCACCGCCCCAAGGAGGGGGTGATGACCGGCGTTCGTTCTTCTTCTTCTGGAATGTTTT 2409

Qy    2375 GAGAATAGTTGTGGGCCATGAAAGTGAATGACAAGAAAGACATTCAAGATTCCTGGCATGT 2434
           ||||||| |||||||||||||||||||||||||||||||||| ||||||||||
Db    2410 GAGAATAGTTGTAGCCCATGAAAGGTTGGCGACAAAGAGGGGAGCATCTCACGGTTCCTGGAGTGT 2469

Qy    2435 AACCCAGTGATGATGATTGATGATGATGGCATCAAGGTTGTCTGCTGCTCACCAGGGCCCGGGTACTTCTGG 2494
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2470 AATCCAGTGATGATGATTGATGATGGCCATCAAAAGTTTCTGCTGCTCACAGGGCCCGGATACTTCTGG 2529

Qy    2495 GGTAACCTACCCGGGAATGAACAGGCCCCGTCAGTGGCTTCAAAGAATGATAAGCTCGGAGCTG 2554
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2530 GGCAACCTACCCGGGAATGGACAGGCCCGTGATAGCATCAAAAGAATGATAAAACTCGAGCTG 2589

Qy    2555 CAGGACTGCCTGGGACGTTCAGTTAGGGACACAGCGAAGTTTAAAGAAAGTTGCAGACAATAACCACC 2614
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2590 CAGGACTGCTTGGAATACAATAGGCAATAGCCAAGTTAAAAGAAAGTTACAGACAATAACCACC 2649

Qy    2615 AAGTCGAAACTCCATCAGACACAGGGCAAAAAACCACTTTCCCTGTAGTCATGATGGCAAG 2674
           ||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db    2650 AAGTCGAAACTCGATCAAAAACAGGGAAAAAACCACTTTTGCCCTGTTGTCATGAATGGCAAA 2709

Qy    2675 GACGACAGCGTTTTTGTGGTGCCACTCTGAAAGCTGAAAAGGATTCTTCGGCTTCCCTGTTCTTGTGAAATGGCAAA 2734
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2710 GAAGATGTTTTTGCTGGTGCACTGAGCTCTGAAAGGATATCTTTGGCTTTTCCTGTGCACTACACA 2769

Qy    2735 GACGTGTCCAAACATGGGCGCGGGCGGCCCGGTATCAGTCCTAC 2794
           |||||||||||||||||||||||||||||||||||||||||||
Db    2770 GACCTGTCCAACATGGGCCGTGGTGCCCGGCCAGAAGCGTGGGAACGGTG 2829

Qy    2795 CCGGTCATCAGACACACCTGTTTGCCCCCTTGAAGGACTACTTTTGCTTGTGAATAGTTCTAC 2854
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2830 CCTGTCATCCGACACACTCTTCCCCCTCTGAAAGGACTACTTTTGCATGTGAATAGTTCCAG 2889

Qy    2855 GACGGTGTCCAACATGGGCGGGGGCGGGCCCGGTCGAGAAGCTGCTGGGGCCAGGTCCCTCCCCTG 2907
           ||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2890 CCAGGGCCCCAAGCCCACTGGGGTGTGTGGCAGGAGCCAGGAGGTGCAAGGTGTGATTCCTG 2949

Qy    2908 AAGGCACCTCACCCTGTCCCCCTTTTTAGCTCACCTGTGTGTGGGGCCTCACACTACCACTGTACCT 2967
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2950 AAGGCA--TCCCAGCCCTGCTCTTCTCTCAGTGTGTTGGACAGCCTTAACACGGTGTACT 3003

Qy    2968 CAGCTTTCTCCCTGCTCAGTGGGAGCAGAGCCCTCCGGCCCTTGCAGGGGAGCCCC--CGGT 3025
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db    3004 CAGTTCCCGTCCTGCTCAGTGGGGGCAGAGCCCACCTGACTCTTGCAGGGGGTAGCCCTGAGGT 3063

Qy    3026 GCTCCCCTCCCGTGTGCACAGCTCAGACCTGCTGCTTAGAGTAGCCCGGGCCATGGGTGCTGCATCAT 3085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3064 GCCGCCTCCTTGTTGCACAAAATCAGACCTGGTCCTTGGAGCAGCCTGCTAACACGGTGCTCAT 3123

Qy    3086 GTTTCTTACCCTGAAACTTTTAAAACTTGAAGTAGCTAGTAAGATGGGGCTTTCTTTTA--- 3142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3124 TTTTTTTTTCTTCTCCTAAAACTTTTAAAAACTTGAAGTAGGGAGGCCTTTTTTTTTTC 3183

Qy    3143 CCCTCCTGAGTTTATCACTC--AGAAGTGATGGCTAAGATACCAAAAAAAACAAAAA 3200
           ||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    3184 CCTTCCTGGGTCTTACCACTCAGAGAACACAAATGGCTAAGAACCACAGTGCCGAC 3243

Qy    3201 CAGGAAACAAAAAAAACAAAAAAAAAACCTCAACAGCTCTTTAGTACTCAAGGTTCATTGCTGCA 3260
           ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3244 AGCTCTCCAATACTCAGGTTAATGCTGAAAAATCATTCCAAAAGACACAGTTAATTGCAAAGAGTTT 3303

Qy    3261 AAATCACTTGAGATTTTGTTTTTTAAGTAAACCCGTGCTCCACATTTGCTTGGAGGATGCTAT 3320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3304 AA------TTTTGAAAAACTGGCTACTGCTCTGTGTTTTACAGACGTGTCCAGT 3350

Qy    3321 TGTGAAATGTGGGCTCAGATGACGCAAGGTCAAGGGGGCCAAAAAAAAATTCCCCTCTCCCCC 3380
```

RESULT 14

LOCUS   AF156488   4145 bp   mRNA   linear   PRI 05-SEP-1999
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)
            mRNA, complete cds.
ACCESSION   AF156488
VERSION   AF156488.1   GI:5823167
KEYWORDS   .
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4145)
  AUTHORS   Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.
            and Li,E.
  TITLE   Cloning, expression and chromosome locations of the human DNMT3
            gene family
  JOURNAL   Gene 236 (1), 87-95 (1999)
  MEDLINE   99365304
   PUBMED   10433969
REFERENCE   2 (bases 1 to 4145)
  AUTHORS   Xie,S. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-JUN-1999), CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
FEATURES   Location/Qualifiers
     source   1..4145
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between D20S182 and D20S106; 39.9-50.2 cM"
     gene   1..4145
                     /gene="DNMT3B"
     CDS   115..2676
                     /gene="DNMT3B"
                     /note="de novo DNA methyltransferase"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3 beta 1"
                     /protein_id="AAD53063.1"
                     /db_xref="GI:5823168"
                     /translation="MKGDTRHLNGEEDAGGREDSILVNGACSDQSSDSPPILEAIRTP
                     EIRGRSSSRLSKREVSSLLSYTQDLTGDGDEDGEDGSDTPVMPKLFRETRTRSESPA
                     VRTRNNNSVSRERHRPSPRSTRGRGRNHVDESPVEFPATRSLRRRATASAGTPWPS
                     PSSYLPIDLTDDTEDTHGTPQSSSTPYARLAQDSQGMESPQVEADSGDGDSSEYQ
                     DGKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKROAMSGMRWVQWFGDGKFSEVSADK
                     LVALGLFSQHFNLATFNKLVSYRKAMYHALEKARVRAGKTFPSSPGSLEDQLKPMLE
                     WAHGGFKPTGIEGLKPNNTQPVVNKSKVRRAGSRKLESRRYENKTRRRTADDSATSDY
                     CPAPKRLKTNCYNNGKDRGDEDQSREOMASDVANNKSSLEDGCLSCGRKNPVSEHPLF
                     EGGLCQTCRDRFLELFYMYDDDGYOSYCTVCCGRERLLCSNTSCCRCFVCELVLV
                     GTGTAAEAKLQDPWSCYMCLPORCHGVLRRRKDNWVRLQAFFISDTGLEYEAPKLYPA
                     IPAARRRPIRVLSLFDGIATGYLLVLKELGIRVGKYVASEVCEESIAVGTVKHEGNIKY
                     VNDVRNITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEYHLLNYS
                     RKEGDDRPFFWMFENVAMKVGDKRDISRFLECNPVMIDAIKVSAAHRARYEWGNLP
                     GMNRPVIASKNDKLELQDCLEYNRIAKLKKVQTITTKSNSIIKQSKNQLFPVVMNGKED
                     VLMCTELERIFGFPVHYTDVSNMGRGAROKLLGRSMSVPVIRHLFAPLKDYFACE"

Query Match   43.2%;   Score 1812.2;   DB 9;   Length 4145;
Best Local Similarity   76.2%;   Pred. No. 0;
Matches 2532;   Conservative 0;   Mismatches 653;   Indels 138;   Gaps 19;

BASE COUNT   1041 a   1083 c   1096 g   925 t
ORIGIN

```
Qy        1241 TCCAGCAGTCCTGGAGAGTCACTGGAGGACCAGCTGAAGCCCATGCTGGAGTGGGCCCAC 1300
                ||||||| |||||||| ||| ||||||||||||||||||||||| ||||||||||||||
Db        1066 CCCAGCAGCCCTGGAGACTCATTGGAGGACCAGCTGAAGCCCATGTTGGAGTGGGCCCAC 1125

Qy        1301 GGTGGCTTCAAGCCTACTGGGATCGAGGGCCTCAAACCCAACAAGAAGCAACCAGTGGTT 1360
               || ||||||||||| ||||||||||||||||||||||||||||| | |||||||||||||
Db        1126 GGGGGCTTCAAGCCCACTGGGATCGAGGGCCTCAAACCCAACAACACGCAACCAGTGGTT 1185

Qy        1361 AATAAGTCGAAGGTGCGTCGTTCAGACAGTAGGAACTTAGAACCCAGGAGACGCGAGAAC 1420
               ||||||||||||||||||||| ||| ||||||||| |||||| | |||| |  |||||||
Db        1186 AATAAGTCGAAGGTGCGTCGTGCAGGCAGTAGGAAATTAGAATCAAGGAAATACGAGAAC 1245

Qy        1421 AAAAGTCGAAGACGCACAACCAATGACTCTGCTGCTTCTG---AGTCCCCCCCACCCAAG 1477
               || | ||||||||||||| |  | ||||| ||  | ||||   | | |||| ||||||||
Db        1246 AAGACTCGAAGACGCACAGCTGACGACTCAGCCACCTCTGACTACTGCCCCGCACCCAAG 1305

Qy        1478 CGCCTCAAGACAAATAGCTAT---GGCGGGAAGGACCGAGGGGAGGATGAGGAGAGCCGA 1534
               ||||||||||||||| |||||     ||| || ||||||||||| || ||  ||||||||
Db        1306 CGCCTCAAGACAAATTGCTATAACAACGGCAAAGACCGAGGGGATGAAGATCAGAGCCGA 1365

Qy        1535 GAACGGATGGCTTCTGAAGTCACCAACAACAAGGGCAATCTGGAAGACCGCTGTTTGTCC 1594
               ||||  |||||||| || ||  ||||||||||| |||  ||||||||  ||||||||||
Db        1366 GAACAAATGGCTTCAGATGTTGCCAACAACAAGAGCAGCCTGGAAGATGGCTGTTTGTCT 1425

Qy        1595 TGTGGAAAGAAGAACCCTGTGTCCTTCCACCCCCTCTTTGAGGGTGGGCTCTGTCAGAGT 1654
               ||||| | ||| ||||| |||||||||||||| ||||||||||| |||||||||||||
Db        1426 TGTGGCAGGAAAAACCCCGTGTCCTTCCACCCTCTCTTTGAGGGGGGGCTCTGTCAGACA 1485

Qy        1655 TGCCGGGATCGCTTCCTAGAGCTCTTCTACATGTATGATGAGGACGGCTATCAGTCCTAC 1714
               ||||||||||||||||| ||||| || |||||||||||||| || ||||||||||| |||
Db        1486 TGCCGGGATCGCTTCCTTGAGCTGTTTTACATGTATGATGACGATGGCTATCAGTCTTAC 1545

Qy        1715 TGCACCGTGTGCTGTGAGGGCCGTGAACTGCTGCTGTGCAGTAACACAAGCTGCTGCAGA 1774
               ||||| |||||||| |||||||| || |||||||| ||||| ||||| ||||||||| |
Db        1546 TGCACTGTGTGCTGCGAGGGCCGAGAGCTGCTGCTTTGCAGCAACACGAGCTGCTGCCGG 1605

Qy        1775 TGCTTCTGTGTGGAGTGTCTGGAGGTGCTGGTGGGCGCAGGCACAGCTGAGGATGCCAAG 1834
               || |||||||||||||| |||||||||||||||||| |||||||||| |  || ||||||
Db        1606 TGTTTCTGTGTGGAGTGCCTGGAGGTGCTGGTGGGCACAGGCACAGCGGCCGAGGCCAAG 1665

Qy        1835 CTGCAGGAACCCTGGAGCTGCTATATGTGCCTCCCTCAGCGCTGCCATGGGGTCCTCCGA 1894
               || ||||| |||||||||||||| ||||| ||||| |||||||| |||||  ||||| ||
Db        1666 CTTCAGGAGCCCTGGAGCTGCTACATGTGTCTCCCGCAGCGCTGTCATGGCGTCCTGCGG 1725

Qy        1895 CGCAGGAAAGATTGGAACATGCGCCTGCAAGACTTCTTCACTACTGATCCTGACCTGGAA 1954
               ||| |||| || |||||| |||||||||| | ||||||||| | |||  | |  || |||
Db        1726 CGCCGGAAGGACTGGAACGTGCGCCTGCAGGCCTTCTTCACCAGTGACACGGGGCTTGAA 1785

Qy        1955 GAATTTGAGCCACCCAAGTTGTACCCAGCAATTCCTGCAGCCAAAAGGAGGCCCATTAGA 2014
                |    ||  | |||||| ||||||| || ||||| ||||||  |||| |||||||| ||
Db        1786 TAC---GAAGCCCCCAAGCTGTACCCTGCCATTCCCGCAGCCCGAAGGCGGCCCATTCGA 1842

Qy        2015 GTCCTGTCTCTGTTTGATGGAATTGCAACGGGGTACTTGGTGCTCAAGGAGTTGGGTATT 2074
               ||||||||  |||||||||| || || || || ||| | || ||||| |||||||| ||
Db        1843 GTCCTGTCATTGTTTGATGGCATCGCGACAGGCTACCTAGTCCTCAAAGAGTTGGGCATA 1902

Qy        2075 AAAGTGGAAAAGTACATTGCCTCCGAAGTCTGTGCAGAGTCCATCGCTGTGGGAACTGTT 2134
               || || | ||||||| | ||  || ||||| ||||  |||||||| |||||  ||||| ||
Db        1903 AAGGTAGGAAAGTACGTCGCTTCTGAAGTGTGTGAGGAGTCCATTGCTGTTGGAACCGTG 1962

Qy        2135 AAGCATGAAGGCCAGATCAAATATGTCAATGACGTCCGGAAAATCACCAAGAAAAATATT 2194
               ||||| || ||  | |||||||| || || |||||  |||| ||||| ||||||||||||
Db        1963 AAGCACGAGGGGAATATCAAATACGTGAACGACGTGAGGAACATCACAAAGAAAAATATT 2022

Qy        2195 GAAGAGTGGGGCCCGTTCGACTTGGTGATTGGTGGAAGCCCATGCAATGATCTCTCTAAC 2254
               ||||| |||||||| || |||||||||||||| |||||||||||||| |||||||| ||
Db        2023 GAAGAATGGGGCCCATTTGACTTGGTGATTGGCGGAAGCCCATGCAACGATCTCTCAAAT 2082

Qy        2255 GTCAATCCTGCCCGCAAAGGTTTATATGAGGGCACAGGAAGGCTCTTCTTCGAGTTTTAC 2314
               || ||||| ||| | |||||  | |||||||| |||||  ||||||||||||| ||||||
Db        2083 GTGAATCCAGCCAGGAAAGGCCTGTATGAGGGTACAGGCCGGCTCTTCTTCGAATTTTAC 2142

Qy        2315 CACTTGCTGAATTATACCCGCCCCAAGGAGGGCGACAACCGTCCATTCTTCTGGATGTTC 2374
               ||| ||||||||||  | |||||||||||||| ||  |||| || ||||||||||||||
Db        2143 CACCTGCTGAATTACTCACGCCCCAAGGAGGGTGATGACCGGCCGTTCTTCTGGATGTTT 2202

Qy        2375 GAGAATGTTGTGGCCATGAAAGTGAATGACAAGAAAGACATCTCAAGATTCCTGGCATGT 2434
               ||||||||||| |||||||| ||    |||||||  ||||||||| | |||||||  |||
Db        2203 GAGAATGTTGTAGCCATGAAGGTTGGCGACAAGAGGGACATCTCACGGTTCCTGGAGTGT 2262

Qy        2435 AACCCAGTGATGATCGATGCCATCAAGGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGG 2494
               || ||||||||||| ||||||||||| || |||||||||||||||||||| |||||||||
Db        2263 AATCCAGTGATGATTGATGCCATCAAAGTTTCTGCTGCTCACAGGGCCCGATACTTCTGG 2322

Qy        2495 GGTAACCTACCCGGAATGAACAGGCCCGTGATGGCTTCAAAGAATGATAAGCTCGAGCTG 2554
               || ||||||||||| ||||||||||||||||| || |||||||||||||| |||||||||
Db        2323 GGCAACCTACCCGGGATGAACAGGCCCGTGATAGCATCAAAGAATGATAAACTCGAGCTG 2382

Qy        2555 CAGGACTGCCTGGAGTTCAGTAGGACAGCAAAGTTAAAGAAAGTGCAGACAATAACCACC 2614
               ||||||||| |||| | || ||||| ||| |||||||||||||| |||||||||||||||
Db        2383 CAGGACTGCTTGGAATACAATAGGATAGCCAAGTTAAAGAAAGTACAGACAATAACCACC 2442

Qy        2615 AAGTCGAACTCCATCAGACAGGGCAAAAACCAGCTTTTCCCTGTAGTCATGAATGGCAAG 2674
               ||||||||||| |||| |||||| |||||||| ||||||||||| ||||||||||||||
Db        2443 AAGTCGAACTCGATCAAACAGGGGAAAAACCAACTTTTCCCTGTTGTCATGAATGGCAAA 2502

Qy        2675 GACGACGTTTTGTGGTGCACTGAGCTCGAAAGGATCTTCGGCTTCCCTGCTCACTACACG 2734
               || || |||||||||||||||||||||||||||||||| ||||| ||||  ||||||||
Db        2503 GAAGATGTTTTGTGGTGCACTGAGCTCGAAAGGATCTTTGGCTTTCCTGTGCACTACACA 2562

Qy        2735 GACGTGTCCAACATGGGCCGCGGCGCCCGTCAGAAGCTGCTGGGCAGGTCCTGGAGTGTA 2794
               |||||||||||||||||||| || ||||| |||||||||||||| ||||||||||| ||
Db        2563 GACGTGTCCAACATGGGCCGTGGTGCCCGCCAGAAGCTGCTGGGAAGGTCCTGGAGCGTG 2622

Qy        2795 CCGGTCATCAGACACCTGTTTGCCCCCTTGAAGGACTACTTTGCCTGTGAATAGTTCTAC 2854
               || |||||| ||||||| || |||||  |||||||||||||||| |||||||||||| |
Db        2623 CCTGTCATCCGACACCTCTTCGCCCCTCTGAAGGACTACTTTGCATGTGAATAGTTCCAG 2682

Qy        2855 CCAGGACTGGGGAGCTCTCG-------GTCAGAGCCAGTGCCCAGAGTCACCCCTCCCTG 2907
               ||||| |    |  | || |       | |||||||||  |||||         | ||||
Db        2683 CCAGGCCCCAAGCCCACTGGGGTGTGTGGCAGAGCCAGGACCCAGGAGGTGTGATTCCTG 2742

Qy        2908 AAGGCACCTCACCTGTCCCCTTTTTAGCTCACCTGTGTGGGGCCTCACATCACTGTACCT 2967
               ||||||  || || | |||   | |  |||| ||||||||| | |     |  |||||||
Db        2743 AAGGCA--TCCCCAGGCCCTGCTCTTCCTCAGCTGTGTGGGTCAT----ACCGTGTACCT 2796

Qy        2968 CAGCTTTCTCCTGCTCAGTGGGAGCAGAGCCTCCTGGCCCTTGCAGGGGAGCCC--CGGT 3025
               ||| |  ||| ||||||||||| |||||||| |||| | ||||||||||   ||   |||
Db        2797 CAGTTCCCTCTTGCTCAGTGGGGGCAGAGCCACCTGACTCTTGCAGGGGTAGCCTGAGGT 2856

Qy        3026 GCTCCCTCCGTGTGCACAGCTCAGACCTGGCTGCTTAGAGTAGCCCGGCATGGTGCTCAT 3085
               ||  ||||| ||||||||  |||||||||||||||| ||| ||||   || |||||||||
Db        2857 GCCGCCTCCTTGTGCACAAATCAGACCTGGCTGCTTGGAGCAGCCTAACACGGTGCTCAT 2916

Qy        3086 GTTCTCTTACCCTGAAACTTTAAAACTTGAAGTAGGTAGTAAGATGGCTTTCTTTTA--- 3142
                || ||||  ||| ||||||||||||||||||||||||| ||  ||||||| ||||
Db        2917 TTTTTCTTCTCCTAAAACTTTAAAACTTGAAGTAGGTAGCAACGTGGCTTTTTTTTTTTC 2976

Qy        3143 CCCTCCTGAGTTTATCACTC--AGAAGTGATGGCTAAGATACCAAAAAAACAAACAAAAA 3200
               || ||||| || || |||||  ||||   ||||||||||||||||||  |||      |
Db        2977 CCTTCCTGGGTCTACCACTCAGAGAAACAATGGCTAAGATACCAAAACCACAGTGCCGAC 3036

Qy        3201 CAGAAACAAAAAACAAAAAAAAACCTCAACAGCTCTCTTAGTACTCAGGTTCATGCTGCA 3260
                     | || |   |    ||  || || |    ||  ||        | || |
Db        3037 AGCTCTCCAATACTCAGGTTAATGCTGAAAAATCATCCAAGACAGTTATTGCAAGAGTTT 3096

Qy        3261 AAATCACTTGAGATTTTGTTTTTAAGTAACCCGTGCTCCACATTTGCTGGAGGATGCTAT 3320
               ||             || ||   || |      ||||     ||| | |  |  |||  |
Db        3097 AA-------------TTTTTGAAAACTGGGTACTGCTATGTGTTTACAGACGTGTGCAGT 3143

Qy        3321 TGTGAATGTGGGCTCAGATGAGCAAGGTCAAGGGGCCAAAAAAAATTCCCCCTCTCCCCC 3380
               |||     ||     | | ||      | | ||  |   |     ||  |||    |
Db        3144 TGTAGGCATGTAGCTACAGGACATTTTTAAGGGCCCAGGATCGTTTTTTCCCAGGGCAAG 3203

Qy        3381 CAGGAGTATTTGAAGATGATGTTTATGGTTTAAGTCTTCCTGGCACCTTCCCCTTGCTTT 3440
               ||| ||     ||| ||| ||| ||||  |    | | || ||||| |||||||||| |
```

```
RESULT 15
AF176228

LOCUS       AF176228        4267 bp    mRNA    linear   PRI 26-OCT-1999
DEFINITION  Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA,
            complete cds.
ACCESSION   AF176228
VERSION     AF176228.1  GI:6118091
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4267)
  AUTHORS   Xu,G.-L., Bestor,T.H., Bourc'his,D., Hsieh,C.-L., Tommerup,N.,
            Bugge,M., Hulten,M., Qu,X., Russo,J.J. and Viegas-Pequignot,E.
  TITLE     Chromosome instability and immunodeficiency syndrome caused by
            mutations in a DNA methyltransferase gene
  JOURNAL   Nature (1999) In press
REFERENCE   2  (bases 1 to 4267)
  AUTHORS   Xu,G.L. and Bestor,T.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) Genetics and Development, Columbia
            University, 701 West 168 St., New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1..4267
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20q11-q13"
     gene            1..4267
                     /gene="DNMT3B"
     CDS             265..2802
                     /gene="DNMT3B"
                     /note="mutated in human chromosome instability and
                     immunodeficiency disease known as ICF syndrome; contains
                     alternative 5' exon 1B"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3B"
                     /protein_id="AAF04015.1"
                     /db_xref="GI:6118092"
                     /translation="MEPSEPEPPSLESMKDTRHLNGEEDAGGREDSILVNGACSDQSS
                     DSPPIEAIRPEIRGRRSSSRLSKREVSSLLSYTQDLTGDGDEDGDGSDPVMPKL
                     FRETRTRSESPAVRTNNNSVSSRERHRPSPRSTRGRQGRNHVDESPVEPATRSLRR
                     RATASAGTPWPSPSSYLTIDLTDDTEDTHGTPQSSSTPYARLAQDSDGGMESPQVE
                     ADSGDGDSSEYQDGKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMSGMRKVQWF
                     GDGKFSEVSADKLVALGLFSQHFNLATFNKLVSRKAMYHALEKARVRAGKTFPSSPG
                     DSLEDQLKPMLEWAHGGFKPTGIEGLKPNNTQPENKTRRTADDSATSDYCPAPKRLK
                     TNCYNNGKDRGDEDQSREQMASDVANNKSSLEDCCLSCGRKNPVSFHPLFEGGLCQTC
                     RDRFLELFYMDDDGYQSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGTGTAAEA
                     KLQEPWSCYMCLPQRCHGVLRRRKDMNVRLQAFFSDTGLEYEAPKLYPAIPAARRP
                     IRVLSLFDGIATGYLVLKELGIKVGKYVASEVCEESIAVGTVHEGNIKYVNDVRNIT
                     KKNIEEWGPPDLVIGGSPCNDLSNVNPARRGLYEGTGRLFFEEYHLLNYSRPKEGDDR
                     PFFWMFENVVAMKVGDKRDISRFLECNPVMIDAIKVSAAHRARYFWGNLPGNRRPVIA
                     SKNDKLELQDCLEYNRIAIKLKKVQTITKSNSIKQGKNQLFPVVMNGKEDVLWCTELE
                     RIFGFPVHYTDVSNMGRGARQKLLGRSWSVPVIRILFAPLKDYFACE"
```

Qy	3204	CAGAAG----AGAAAATGTTGTATATGTCT----TTTACCGGCGACATTCCCCTTGCCTA 3255
		||||| ||||||| |||||||||
Db	3441	GGTACAAAGGCGCTGAAGTCCGTGTTGG------TCTTGTGTAGCCATTTCCCAGGATGATGATG 3493
Qy		
Db	3256	AATACAAAGGCGGCTGGAGCTCTGCGACGGGACCTTATTAGAGTATTTTCCACAATGATGATGATT 3315
		||||||||||||||||| || ||||
Qy	3494	TCAGCAGGGATGACATCACCACC 3516
		|||||||||||||||| |||| |
Db	3316	TCAGCAGGGATGACGTCATCATC 3338

```
BASE COUNT   1069 a   1098 c   1147 g   953 t
ORIGIN
```

Qy	233	CTTGGGACCAAATCCAAGGGCCCTTCTTTC·AGGAAACAATGAAGGAGGAAGAGAGACATCT 291
Db	264	CATGGGAACCAAAGTCCTGAGCCCTCCAAGCTTGGAAAACATGAAGGAGGACACCAGGCATCT 323
Qy	292	GAATGAAGAACAGAGGGGTGCCAACGGCGTATGAGGAGTGCATTATCGTTAATGGGAAACTTCAG 351
Db	324	CAATGGAGAGGAGGAGCAGCCCGGCGGGAGGGAAGACTCGGAATCCGTCGTCAACGGGCCTGCAG 383
Qy	352	TGACCAGTCCTCAGACACGAAGGATGCTCCCTACCCCAGTCTTGGAGGCAATCTGCAC 411
Db	384	CGACCAGTCTCTCCGA----CTCCGCCCCAAATCCTGGAGGCTA- 421
Qy	412	AGAGCCAGTCTTGCACACCACAGAGACACAGAGGAGGACGCCGGGTGAAAGTCTCCCGCTGTCTTAAGAG 471
Db	422	-----TCCCCACCCCGGAGGGATCCAGGAGGGCGAAGCATCAAAGCTGCGGCAGGACTCTTCCAAGAG 473
Qy	472	GGAGGTCTCCAGCCTTCTGAATTTACACGCACAGGAGACAATGACAGGGAGATGGAGACAAGATGA 531
Db	474	GGAGGTGTCCAGTCTCCTTAGCCTACACACAGGACTTGACAGGGCCGATGGCGAC------- 525
Qy	532	TCGAAGTAGATGAATGGGAAATGGCTCTGATTATTC----TAATGCCAAAGCTCACCCGGTGAGAC 588
Db	526	----GGGGAAGAGATGGGGGATGGCTCTGCAACACCCCACTCATGCCAAAGCTCTTCG----- 575
Qy	589	CAAGGACACCAGGACGCCGCGCTCGTGAAAGACCCGGCGGTGCTCCAAACCCGGACATAGCAATGGGAC 648
Db	576	---GGAAACCAGGACTCTGTTCAGAAAGCCCACCTGTCCGGACTCGAAAATAACAACAGTGT 632
Qy	649	CTCCAGCTTGGAGAGCGCAAAAGACCCTCCCCCAGAATCACCCGAGCTCGGGCGAGGGCCA 708
Db	633	CTCCAGCCGCGGGAGGGCACACAGCCTTCCCACGTTCCACCGGAGCCCGGCCAGGCGGGCCAA 692
Qy	709	CCAATGTCGCAGGAGTACCCTGTCGGAGTTCCGGGCTACCAGGTCTCGGAGACGTCGAGCATC 768
Db	693	CCAATGTGGACGAGTCCCCCGTCGGAGTTCCCGGGCTACCAGGTCCCTGAGACGGGCGGGCAAC 752
Qy	769	GTCTTCAGGCAAGCACCCCCATGGTCATCCCCGTCCCCAGCGTGGACTTCA---- 815
Db	753	AGCATCGGCAGGAGAACGCCATGGCCGTCCCCTCCCAGCTCTTACCTTACCATCAGACCTCAC 812
Qy	816	----TGGAAGAAGTGACAACTTAAGACGGTCAGTACCCCATCAGTTGACTT 861
Db	813	AGAGCGACACAGAGGACACACATGGGACGGCCCCAGAGCCAGCAGTACCCCCTACGGCCCGCCT 872
Qy	862	GAGCCAGGATGGAGAATCAGGAGGGTATGGATAACCACAGTGGATGCAGAAGGACCAGAGA 921
Db	873	AGCCCCAGGACACAGCCAGCAGGGGGGCCATGGAGTCCCCGCAGGTGGAGGCAGACAGTGGAGA 932
Qy	922	TGGGAGACAGCACACAGAGTATCAGGATGATAAAGAGTTTGGAGATGAGCCTGGTGGGG 981
Db	933	TGGGAGACAGTTCAGAGTAGTCAGGAGTGGGAAGGAGTTTGGAATAGGGGAACCTGGTGGGG 992
Qy	982	AAAGATCAAAGGCGGTTCTCCCTGTGGCCTGCCATAGTGGTGGTGTCCCTGGAAAGCCACCTCCAA 1041
Db	993	AAAGATCAAAGGGCGTTCCTCCCTGGTGGCCCGCCCATAGTGGTTGGAAGGCCACCTCCAA 1052
Qy	1042	GCCACAGGCCCCATCCCCGGAAATGCGCTAGACGATGGTGTTGGTGATGGCAAGTTTTCTGA 1101
Db	1053	GCCACAGCCTATCTCTGGCATCCCGGTGGGTCCCAGTGGTTCCGGATGGCAAGTTCTCCGA 1112
Qy	1102	GATCTCTGCTGACAAACTGGTGGCTCTCGGGGTGTTCAGCCCAGCACTTTAATCTGGCTAC 1161
Db	1113	GGTCTCTGCAGACAAACTGGTGGCCACTGGGCTGTTCAGCCAGCACTTTAATTTGGCCAC 1172
Qy	1162	CTTCAAATAAGCTCGTTTCTTTATAGGAAGGCCACCATGTACCACACTCTGGAGAAAGCCAGGGT 1221
Db	1173	CTTCAAATAAGCTCGTCTCCTATCGAAAAGCCATGTACCACACTCTGGAGAAAGCTAGGGT 1232
Qy	1222	TCGGAGCTGGCAAGAACCTTCTCCAGCAGACCTTCACTGGAGACCAAGCCGAAAGCC 1281
Db	1233	GCGAGCTGGCAAGAACCTTCCCCAGCAGCCCTCGGAGACTCATTGGAGGACCAAGCTGAAAGCC 1292
Qy	1282	CATGCTGGAGTGGGCCCACGGTGCTTCAAGCCTACTGGGGATCAGAGGGGCCTCAAAACCCAA 1341

```
Qy   3475  ATTTCCCAGGATGATGATGAGTGTCAGCAGGGATGACATCACCACC  3516
            | | |||||||||| ||||||||||||||| ||| |--|
Db   3423  TTCCACAATGATGATGATTTCAGCAGGGATGACGTCATCATC  3464
```

Search completed: July 17, 2003, 07:37:39
Job time : 7065.06 secs

RESULT 1
AF068626
LOCUS   AF068626   4195 bp   mRNA   linear   ROD 06-DEC-1999
DEFINITION   Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,
               alternatively spliced, complete cds.
ACCESSION   AF068626
VERSION   AF068626.2 GI:6449469
KEYWORDS   .
SOURCE   Mus musculus.
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4195)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE   Cloning and characterization of a family of novel mammalian DNA

THIS PAGE BLANK (USPTO)

```
JOURNAL      (cytosine-5) methyltransferases
             Nat. Genet. 19 (3), 219-220 (1998)
MEDLINE      98324766
PUBMED       9662389
REFERENCE    2 (bases 1 to 4195)
AUTHORS      Xie,S., Okano,M. and Li,E.
TITLE        Direct Submission
JOURNAL      Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
             Charlestown, MA 02129, USA
REFERENCE    3 (bases 1 to 4195)
AUTHORS      Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
TITLE        Direct Submission
JOURNAL      Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th street,
             Charlestown, MA 02129, USA
REMARK       Sequence update by submitter
COMMENT      On Nov 18, 1999 this sequence version replaced gi:3327979.
FEATURES              Location/Qualifiers
     source           1..4195
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /chromosome="2"
                      /map="84.0 cM"
     gene             1..4195
                      /gene="Dnmt3b"
     misc_feature     1..4195
                      /gene="Dnmt3b"
                      /note="similar to EST sequences deposited in GenBank
                      Accession Numbers AA116694, AA119979, AA177277, AA210568,
     CDS              AA407106, and AA575617"
                      269..2848
                      /gene="Dnmt3b"
                      /function="de novo DNA methylation"
                      /note="alternatively spliced product; contains Cys-rich
                      region; C-terminal region is similar to corresponding
                      region of DNA cytosine-5 methyltransferase 3A"
                      /codon_start=1
                      /product="DNA cytosine-5 methyltransferase 3B1"
                      /protein_id="AAC40178.2"
                      /db_xref="GI:6449470"
                      /translation="MKGDSRHLNEEGASGYEECIIVNGNFSDQSSDIKDAPSPVLE
                      AICTEPVCTPETGRSSSRLSKREVSSLLNYTQDMTGDGRDDEVDDGNGSDIMPK
                      LTRETKDPRTRSESPAVRIRHSNGTSSLERQRASPRITRGRQGRHHVQEYPVEPPATR
                      SRRRASSSASTPKSSPASVDFMEEVTPKSVSTPSVDLSQDGPQEGMDTQVDAESRD
                      GDSTEYQDDKEPGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMRWVQWFGDGKF
                      SEISADKLVALGLFSQHFNLATFNKLVSYRKAMYHTLEKARVRAGKTPFSSSPGESLED
                      QLKPMLEWAHGGEKPTGIEGLKPNKKQPVVNKSKVRRSDSRNLEPRRENKSRRRTTN
                      DSAASESPPKRLKTNSYGGKDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVS
                      FHPLFEGGLCQSCRDRFLELFYMYDEDGYQSYCTVCCEGRELLLCSNTSCCRCFCVEC
                      LEVLVGAGTAEDAKLQEPWSCYMCLPQRCHGVLRRRKDWNMRLQDFFTTDPLEEFEP
                      PKLYPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKH
                      EGQIKYVNDVRKITRKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFY
                      HLLNYTRPKEGDNRPFFWMFENVVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARY
                      FWGNLPGMNRPVWASKNDKLELQDCLEFSRTAKLKKVQTITPKSNSTRQGKNQLFPVV
                      MNGKDDVLWCTELERIFGFPAHYTDVSNMGRGAROKLLGRSWSVPVIRHLFAPLKDYF
                      ACE"
BASE COUNT      1061 a      1059 c      1125 g       950 t
ORIGIN
```

```
Query Match    100.0%;  Score 4195;  DB 10;  Length 4195;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 4195; Conservative  0; Mismatches  0;  Indels  0;  Gaps  0;

Qy         1 GAATTCCGGCGCGCGCGCGGGTTAAGCGCCCAAGTAAACGTAGCGCAGCGATCGGCGCGA 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GAATTCCGGCGCGCGCGCGGGTTAAGCGCCCAAGTAAACGTAGCGCAGCGATCGGCGCGA 60

Qy        61 GATTCGCGGAACCCGACACTCCCGCGCCGCCGCGCGCCGGCCAGGACCCGGCGATCGGGC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GATTCGCGGAACCCGACACTCCCGCGCCGCCGCGCGCCGGCCAGGACCCGGCGATCGGGC 120

Qy       121 GCCCGCGCTACAGCCACCTACGACACGCCCGCCGTCAGGACGCCTTGTGCCCAGACCTTGGAAAACC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GCCCGCGCTACAGCCACCTACGACACGCCCGCCGTCAGGACGCCTTGTGCCCAGACCTTGGAAAACC 180

Qy       181 TCAGGTATATAACCTTTTCCGAGACGCGGGGGGATCTCCCCTCCCCCATCCATAGTGCCCTTGGGAC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TCAGGTATATAACCTTTTCCGAGACGCGGGGGGATCTCCCCTCCCCCATCCATAGTGCCTTGGGAC 240

Qy       241 CAAATCCAGGGCGCCTTCTTCTTCCAGGAAACAAATAAAGGGGAGCACAGAGACATCTGAATGAAGA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAAATCCAGGGCCTTCTTCTTCCAGGAAACAAATGAAGGGGAGACACAGAGACATCTGAATGAAGA 300

Qy       301 AGAGGGGTGCCAGCCGGGTATGAGGAGTGCATTATGTGTTAATCGTTAATGGGAACTCAGTGACCAGTC 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AGAGGGGTGCCAGCCGGGTATGAGGAGTGCATTATGTGTTAATCGTTAATGGGAACTTCAGTGACCAGTC 360

Qy       361 CTCAGACACGAAGGATGCTCCCTCACCCCCAGTCTTGGAGCCAATCTGCACAGAGCCAGT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CTCAGACACGAAGGATGCTCCCTCACCCCCAGTCTTGGAGCCAATCTGCACAGAGCCAGT 420

Qy       421 CTGCACACCAGAGACCAGAGGCCCCCAGGTCAAGCTCCCCGGCTGTCTAAGAGGGAGCTC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTGCACACCAGAGACCAGAGGCCCCCAGGTCAAGCTCCCCGGCTGTCTAAGAGGGAGCTC 480

Qy       481 CAGCCCTTCTGAATTACACGCAGAGACATGACAGAGAGGATGGAGAGCAGAGCAGAGTAGA 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CAGCCCTTCTGAATTACACGCAGAGACATGACAGAGAGGATGGAGAGCAGAGCAGAGTAGA 540

Qy       541 TGATGGGAANTGCTCGAATATTCTAATGCAAAAGCTCACCGTGAGACCAAGGACACCAG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TGATGGGAANTGCTCGAATATTCTAATGCAAAAGCTCACCGTGAGACCAAGGACACCAG 600

Qy       601 GACGCGCTCTGAAAAGCCGGCTGTCCGAACCCCAACATAGCAATGGGACTTCCAGCTTGGA 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 GACGCGCTCTGAAAAGCCGGCTGTCCGAACCCCAACATAGCAATGGGACTTCCAGCTTGGA 660

Qy       661 GAGGCAAAGAAGCCTCCCCCAGAATCACCCGAGGTCGGGCAGGGCGCCACCAGTGTCGCAGGA 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GAGGCAAAGAAGCCTCCCCCAGAATCACCCGAGGTCGGGCAGGGCGCCACCAGTGTCGCAGGA 720

Qy       721 GTACCCTGTGGAGTTTCCGGCGTACCCAGGTCTCGGGACGTCGAGCATCGTTCAGCAAG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GTACCCTGTGGAGTTTCCGGCGTACCCAGGTCTCGGGACGTCGAGCATCGTTCTTCAGCAAG 780

Qy       781 CACGCCATGGTCATCCCCTCCCCAGCCGTCGACTTCATGGAAGAAGTGACACTTAAGAGCGT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACGCCATGGTCATCCCCTCCCCAGCCGTCGACTTCATGGAAGAAGTGACACTTAAGAGCGT 840

Qy       841 CAGTACCCCATCAGTTGACTTGAGCCAGGATGGGAGAGATCAGGATGAGATGAAGACCACA 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CAGTACCCCATCAGTTGACTTGAGCCAGGATGGGAGAGATCAGGAGGGTATGGATACCACA 900

Qy       901 GGTGGAGTGCAGAGAGCAGAGATGGAGAGCAGCAGGATGGAGAGTATCAGGATAAAGAGTTTGG 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GGTGGAGTGCAGAGAGCAGAGATGGAGAGCAGCAGGATGGAGAGTATCAGGATAAAGAGTTTGG 960

Qy       961 AATAGGTGACCTCGTGTGGGGAAGATCAAGGGCTTCTCCTGGTGGCCTGCATGGTGGT 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AATAGGTGACCTCGTGTGGGGAAGATCAAGGGCTTCTCCTGGTGGCCTGCATGGTGGT 1020

Qy      1021 GTCCTGGAAAGCCACCTCCAAGCGACAGGCCATGCCGGCTGGTACAGTGGTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 GTCCTGGAAAGCCACCTCCAAGCGACAGGCCATGCCGGCTGGTACAGTGGTT 1080

Qy      1081 TGGTGATGGCAAGTTTTCTGAGATCTCTGCTGACAAAACTGGTTGGCTCTGGGCTGTTCAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TGGTGATGGCAAGTTTTCTGAGATCTCTGCTGACAAAACTGGTTGGCTCTGGGCTGTTCAG 1140

Qy      1141 CCACCACTTTAATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTACCA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CCACCACTTTAATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTACCA 1200

Qy      1201 CACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGAACCTTCTCCAGCAGTCCTGGAGAGTC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGAACCTTCTCCAGCAGTCCTGGAGAGTC 1260
```

THIS PAGE BLANK (USPTO)

```
Qy   1261 ACTGGAGGACCAGCTGAAGCCTCAGCTGCTGGAGTGGGGCCCACGGGTGGGGCGTTCAAGCCTACTGG 1320
          ||||| ||||||||||||||||||||
Db   1261 ACTGGAGGACCAGCTGAAGCCCATGCTGGAGTGGGGCCCACGGGTGGGGCGTTCAAGCCTACTGG 1320

Qy   1321 GATCGAGGGGCCTCAAACCCAACAAACAAAGAAGCAACCAGTGGTTAATAAGTCGAAGGTGCGTCG 1380
          ||||||
Db   1321 GATCGAGGGGCCTCAAACCCAACAAACAAAGAAGCAACCAGTGGTTAATAAGTCGAAGGTGCGTCG 1380

Qy   1381 TTCAGACAGTAGGAACTTAGAACCCAGGAGACGCGGAACACAAAGTGGAAGACGCACAAC 1440
          |||||||
Db   1381 TTCAGACAGTAGGAACTTAGAACCCAGGAGACGCGGAACACAAAGTGGAAGACGCACAAC 1440

Qy   1441 CAATGACTCTGCTGTCTGCTTCTGAGTCCCCCACCCAGCGCTCAAGACAAATAGCTATGG 1500
          |||||||
Db   1441 CAATGACTCTGCTGTCTGCTTCTGAGTCCCCCACCCAGCGCTCAAGACAAATAGCTATGG 1500

Qy   1501 CGGGAAGGAGGACCGAGGGGAGGGATGGGCTTCTGAAGTCACCAA 1560
          |||||||
Db   1501 CGGGAAGGAGGACCGAGGGGAGGGATGGGCTTCTGAAGTCACCAA 1560

Qy   1561 CAACAAGGGCAATCTGGAAGACCGCGTGTTTGTCCTGTGGAAAGAAGAACCCTGTGTGTCCTT 1620
          |||||||
Db   1561 CAACAAGGGCAATCTGGAAGACCGCGTGTTTGTCCTGTGGAAAGAAGAACCCTGTGTGTCCTT 1620

Qy   1621 CCACCCCCTCTCTTTGAGGGTGGGGCTCTGTGTCAGGATTGCCCGGGGACCCTAGAGCTCTT 1680
          |||||||
Db   1621 CCACCCCCTCTCTTTGAGGGTGGGGCTCTGTGTCAGGATTGCCCGGGGACCCTAGAGCTCTT 1680

Qy   1681 CTACATGTATGATGAGGAGGACCTATCAGTCCTCTACTGCACCGTGTCGTCTGCGTGTCCTT 1740
          |||||||
Db   1681 CTACATGTATGATGAGGAGGACCTATCAGTCCTCTACTGCACCGTGTCGTCTGCGTGTCCTGA 1740

Qy   1741 ACTGCTGCTGTGTGAGTAACACAACGCTGCTGCAGATCTGCTTCTGTGTGGACTGTGGAGGT 1800
          |||||||
Db   1741 ACTGCTGCTGTGTGAGTAACACAACGCTGCTGCAGATCTGCTTCTGTGTGGACTGTGGAGGT 1800

Qy   1801 GCTGGTGGGGCGGCAGGACACAGCTGAGGATGCCAGAGCTGCCAGTGGAAGCTGCTATAT 1860
          |||||||
Db   1801 GCTGGTGGGGCGGCAGGACACAGCTGAGGATGCCAGAGCTGCCAGTGGAAGCTGCTATAT 1860

Qy   1861 GTGCCTCCCTCAGGCGCTGCCATGGGGGTCCTCCGACGCGAGGAAAGAGATTGGAATCGCGCCT 1920
          |||||||
Db   1861 GTGCCTCCCTCAGGCGCTGCCATGGGGGTCCTCCGACGCGAGGAAAGAGATTGGAATCGCGCCT 1920

Qy   1921 GCAAGACTTCTTCACTACTACTGAATCCTGACCTGGAAGAATTGACCCACCAAGTTGTACCC 1980
          |||||||
Db   1921 GCAAGACTTCTTCACTACTACTGAATCCTGACCTGGAAGAATTGACCCACCAAGTTGTACCC 1980

Qy   1981 AGCAAATTCCTGCACCAAAAGGAGGGCCCATTAGAGTCTGCTGTCTGTTTGCATGGAATTGC 2040
          |||||||
Db   1981 AGCAAATTCCTGCACCAAAAGGAGGGCCCATTAGAGTCTGCTGTCTGTTTGCATGGAATTGC 2040

Qy   2041 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTACATTGCCTCCGA 2100
          |||||||
Db   2041 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTACATTGCCTCCGA 2100

Qy   2101 AGTCTGTGCAGAGTCCATCGCTGTGTGGGAACTGTTTAAGCATGAAGGCCCATTTAATGT 2160
          |||||||
Db   2101 AGTCTGTGCAGAGTCCATCGCTGTGTGGGAACTGTTTAAGCATGAAGGCCCATTTAATGT 2160

Qy   2161 CAATGACGTCCGGAAAATCACCAAGAAAAATATTGAAGAGTGGGGGCCCCGTTGACTTGGT 2220
          |||||||
Db   2161 CAATGACGTCCGGAAAATCACCAAGAAAAATATTGAAGAGTGGGGGCCCCGTTGACTTGGT 2220

Qy   2221 GATTGGTGGTGGAAGCCCATGCAATGAGATCTCTCAACCTGCCCGCAAAGGTTTATA 2280
          |||||||
Db   2221 GATTGGTGGTGGAAGCCCATGCAATGAGATCTCTCAACCTGCCCGCAAAGGTTTATA 2280

Qy   2281 TGAGGGCACAGGAAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCGGCCCCAA 2340
          |||||||
Db   2281 TGAGGGCACAGGAAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCGGCCCCAA 2340

Qy   2341 GGAGGGCGACAACCGGTCCATTCTTCTCTGGATGTGTTCGGCCATGAAAAGTGAA 2400
          |||||||
Db   2341 GGAGGGCGACAACCGGTCCATTCTTCTCTGGATGTGTTCGGCCATGAAAAGTGAA 2400
```

```
Db   2341 GGAGGGCGACAACCGTCCAATTCTTCTTCTGGATGTTCGGCCATGATGATCGAGAAATGTTGTGGCCATGAAAAGTGAA 2400
          |||||||
Qy   2401 TGACAAGAAAGACATCTCAAGAATCCTGGCCATGAATGAACGCCATCAA 2460
          |||||||
Db   2401 TGACAAGAAAGACATCTCAAGAATCCTGGCCATGAATGAACGCCATCAA 2460

Qy   2461 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCCGGGGAAATGAACAGGCC 2520
          |||||||
Db   2461 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCCGGGGAAATGAACAGGCC 2520

Qy   2521 CGTGATGGCTTCAAAGAAGAATGATAAGCTCGAGGCTGCAGGGACTGGGAGTTCAGTAGGAC 2580
          |||||||
Db   2521 CGTGATGGCTTCAAAGAATGATAAGCTCGAGGCTGCAGGGACTGGGAGTTCAGTAGGAC 2580

Qy   2581 AGCAAAGTTAAAGAAAGTGCAGCAAATAACCACCAAGTCGAACTCCATCAGGAGGGCAA 2640
          |||||||
Db   2581 AGCAAAGTTAAAGAAAGTGCAGCAAATAACCACCAAGTCGAACTCCATCAGGAGGGCAA 2640

Qy   2641 AAACCAGCTTTTCCCTGTAGTCAATGAAATGGGCAAGGAACGTTTTGTGGGTGCACTGGAGCT 2700
          |||||||
Db   2641 AAACCAGCTTTTCCCTGTAGTCAATGAAATGGGCAAGGAACGTTTTGTGGGTGCACTGGAGCT 2700

Qy   2701 CGAAAGGAGGATCTTCGGGCTTCCCTGCTCACTCACACGGACGGGGAGGCGCCCGGCGC 2760
          |||||||
Db   2701 CGAAAGGAGATCTACTTTGCCTCGTGAATAGTTCTACCCAGAGCTCGGGGGCGCCCGGCGC 2760

Qy   2821 CTTTGAAGGAGACTACTTTGCCTCTGCTCACTCTCCTCTGCTTGGGGGAGCGGGAGCT 2880
          |||||||
Db   2821 CTTTGAAGGAGACTACTTTGCCTCTGCTCACTCTCCTCTGCTTGGGGGAGCGGGAGCT 2880

Qy   2881 CCGTCAGGAGCTGCTGGGGAGCAGTCACCCCTCCCTGAAAGGCCACCTCACCTGTCCCC 2940
          |||||||
Db   2881 GCCAGTGCCCAGAGTCACCCCTCCCTGAAAGGCCACCTCACCTGTCCCC 2940

Qy   2941 TGTGTGGGGCCTCACATCACTGTACCCTCAGCTTTCTCTGCTCAGTGGGACAGAGCCCTC 3000
          |||||||
Db   2941 TGTGTGGGGCCTCACATCACTGTACCCTCAGCTTTCTCTGCTCAGTGGGACAGAGCCCTC 3000

Qy   3001 CTGGCCCCTTGCACGAGTCACCCCGTGCGGCCCCGGTGGCACAGCTCAGACCTGGCTGCT 3060
          |||||||
Db   3001 CTGGCCCCTTGCACGAGTCACCCCGTGCTGCCCCGGTGGCACAGCTCAGACCTGGCT 3060

Qy   3061 TAGAGTAGGCCCGGCCATGGTCGCTCATGTTCTCTTCTTACCCTGAAACTTTTAAAACTTGAAGTAG 3120
          |||||||
Db   3061 TAGAGTAGGCCCGGCCATGGTCGCTCATGTTCTCTTCTTACCCTGAAACTTTTAAAACTTGAAGTAG 3120

Qy   3121 GTAGTAAGATGGCTTTCTTTTTCTTTTACCCTCCGAGTTTATCACTCAGAAGTGATGGCTAAGAT 3180
          |||||||
Db   3121 CTGGCCCTTGCAGGGGGAGCCCGTGTTGCACAGCTCAGACCTGGCT 3180

Qy   3181 ACCAAAAAAAACAGAAACACAGAAACAGAGCTAGCTCTCTTA 3240
          |||||||
Db   3181 ACCAAAAAAAACAGAAAACACAAACAGAGCTAGCTCTCTTA 3240

Qy   3241 GTACTCAGGTTCATGCTGCAAAATCACTGAGATTTTGTTTGTTTTTTAAGTAACCGTGCTCCA 3300
          |||||||
Db   3241 GTACTCAGGTTCATGCTGCAAAATCACTGAGATTTTGTTTGTTTTTTAAGTAACCGTGCTCCA 3300

Qy   3301 CATTTGCTGAGGAGGATGCTATTTGTGAAGTCAGAAGTGAGCAAGGTCAAGGGGCAAA 3360
          |||||||
Db   3301 CATTTGCTGAGGAGGATGCTATTTGTGAAGTCAGAAGTGAGCAAGGTCAAGGGGCAAA 3360

Qy   3361 AAAAATTCCCCTCTCCCCCCAGGAGGTATTTGAAGATGATGTTTATGGTTTAAGTCTTCC 3420
          |||||||
Db   3361 AAAAATTCCCCTCTCCCCCCAGGAGGTATTTGAAGATGATGTTTATGGTTTAAGTCTTCC 3420

Qy   3421 TGGCAACCTTCCCCTTGCTTTGGTTACAACGGGCCTGAAGTCCTGTTCGTCTTGTAGCCATTTCC 3480
          |||||||
Db   3421 TGGCAACCTTCCCCTTGCTTTGGTTACAACGGGCCTGAAGTCCTGTTCGTCTTGTAGCCATTTCC 3480
```

```
Db     3421 TGGCACCTTCCCCTTGCTTTGGTACAAGGGCTGAAGTCCTGTTGGTCTTGTAGCATTTCC 3480

Qy     3481 CAGGATGATGATGTCAGCAGGGATGACATCACCACCTTTAGGGCTTTTCCCTGGCAGGGG 3540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3481 CAGGATGATGATGTCAGCAGGGATGACATCACCACCTTTAGGGCTTTTCCCTGGCAGGGG 3540

Qy     3541 CCCATGTGGCTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAAGGGTGGG 3600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3541 CCCATGTGGCTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAAGGGTGGG 3600

Qy     3601 TGGAGTGGCCCTCTTCCAGGTGTGAGGGATACGAAGGAGGAAGCTTAGGGAAATCCATTC 3660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3601 TGGAGTGGCCCTCTTCCAGGTGTGAGGGATACGAAGGAGGAAGCTTAGGGAAATCCATTC 3660

Qy     3661 CCCACTCCCTCTTGCCAAATGAGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCC 3720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3661 CCCACTCCCTCTTGCCAAATGAGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCC 3720

Qy     3721 CTAGTTCCTCATGAGAAGCTAGGACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTT 3780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3721 CTAGTTCCTCATGAGAAGCTAGGACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTT 3780

Qy     3781 TGGGGAACTACAGTGAAAACCTTCTGGAGATGTTAAAAGCTTTTTACCCCACGATAGATT 3840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3781 TGGGGAACTACAGTGAAAACCTTCTGGAGATGTTAAAAGCTTTTTACCCCACGATAGATT 3840

Qy     3841 GTGTTTTTAAGGGGTGCTTTTTTTAGGGGCATCACTGGAGATAAGAAAGCTGCATTTCAG 3900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3841 GTGTTTTTAAGGGGTGCTTTTTTTAGGGGCATCACTGGAGATAAGAAAGCTGCATTTCAG 3900

Qy     3901 AAATGCCATCGTAATGGTTTTTAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGA 3960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3901 AAATGCCATCGTAATGGTTTTTAAACACCTTTTACCTAATTACAGGTGCTATTTTATAGA 3960

Qy     3961 AGCAGACAACACTTCTTTTTATGACTCTCAGACTTCTATTTTCATGTTACCATTTTTTTT 4020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3961 AGCAGACAACACTTCTTTTTATGACTCTCAGACTTCTATTTTCATGTTACCATTTTTTTT 4020

Qy     4021 GTAACTCGCAAGGTGTGGGCTTTTGTAACTTCACAGGTGTGGGGAGAGACTGCCTTGTTT 4080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4021 GTAACTCGCAAGGTGTGGGCTTTTGTAACTTCACAGGTGTGGGGAGAGACTGCCTTGTTT 4080

Qy     4081 CAACAGTTTGTCTCCACTGGTTTCTAATTTTTAGGTGCAAAGATGACAGATGCCCAGAGT 4140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4081 CAACAGTTTGTCTCCACTGGTTTCTAATTTTTAGGTGCAAAGATGACAGATGCCCAGAGT 4140

Qy     4141 TTACCTTTCTGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAA 4195
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4141 TTACCTTTCTGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAA 4195
```

RESULT 2
AF151969
LOCUS   AF151969
DEFINITION Mus musculus nu
                nt3h1

THIS PAGE BLANK (USPTO)

4493 TGGTTGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAA 4338

RESULT 3
LOCUS   AF068627

AF068627
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF068627
VERSION   AF068627.2 GI:6449471
KEYWORDS   .
SOURCE   Mus musculus
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4135)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE   Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
  JOURNAL   Nat. Genet. 19 (3), 219-220 (1998)
  MEDLINE   98324766
  PUBMED   9662389
REFERENCE   2 (bases 1 to 4135)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REFERENCE   3 (bases 1 to 4135)
  AUTHORS   Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK   Sequence update by submitter
COMMENT   On Nov 18, 1999 this sequence version replaced gi:3327981.
FEATURES   Location/Qualifiers
     source   1..4135
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="84.0 cM"
     gene   1..4135
                     /gene="Dnmt3b"
     misc_feature   1..4135

4135 bp   mRNA   linear   ROD 06-DEC-1999

THIS PAGE BLANK (USPTO)

CDS   269..2788
        /gene="Dnmt3b"
        /note="similar to EST sequences deposited in GenBank
        Accession Numbers AA116694, AA119979, AA177277, AA210568,
        AA407106, and AA575617"
        /gene="Dnmt3b"
        /function="de novo DNA methylation"
        /note="alternatively spliced product; contains Cys-rich
        region; C-terminal region is similar to corresponding
        region of DNA cytosine-5 methyltransferase 3A"
        /codon_start=1
        /product="DNA cytosine-5 methyltransferase 3B2"
        /protein_id="AAC40179.2"
        /db_xref="GI:6449472"
        /translation="MKGDSRHLNEEGASGYEECIIVNGNFNFSDQSSDPIKDAPSPPVLE
        AICTEPVCTPETPGRSRSSRLSKREVSSLLNYTQDMTGDGDGDRDDEVDDGNGSDLLMPK
        LTRETKDRTRTSRSSPAVRRTRHSNGTSSLERQRASPRITRGRQGRHHVQEYPVEPPATR
        SRRRASSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDQGEMDTTQVDAESRD
        GDSTEYQDDKEFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMPGMMVVOWFDGKF
        SEISADKLVALGLFSQHFNLATFNKLVSYRRAMYHTLEKARVRAGKTPSSSPGESLED
        QLKPMLEWAHGGFKPTGIEGLKPNKQPENSRRTNDSAASESPPRRLKTNSYGG
        KDRGEDESRERNASEVTNNKGNLEDRCLSCGGKKNPVSFHPLFEGGLCQSCRDRFLEI.
        FYMYDEDGYQSYCTVCCEGRELLCSNTSCCRFCVECLEVLVGAGTAEDALCDEPWS
        CYMCLPQRCHGVLRRKDWNMRLQQDFFTDPLEEFPPKLYPAIPAAKRRPIRVLSL
        FDGIATGYLVLKELGIKVEKYIASEVCAEAIAVGTYKHEGOIKYVNDVRKITKKNIEE
        WGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYRLLNYTRPKEGDNRPFFWM
        ENVVAMKVNDKKDISRFLACNPVMIDAIKVSAAHRARYFWGNLPGMNRPVMASKNDKL
        ELQDCLEFSRTAKLKKVQTITTKSNSIROGKNQLFPVVMNGKDDVLWCTELERIGFP
        AHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"

BASE COUNT   1043 a   1048 c   1106 g   938 t

ORIGIN

Query Match   98.9%;   Score 4065;   DB 10;   Length 4135;
Best Local Similarity   98.6%;   Pred. NO. 0;
Matches 4135; Conservative 0; Mismatches 0; Indels 60; Gaps 1;

[sequence alignment blocks follow — paired Qy (query) and Db (database) lines of nucleotide sequence from positions 1 through 1680, with alignment bars between them]

THIS PAGE BLANK (USPTO)

```
Db       1561 CCACCCCCTCTTTGAGGGTGGGCTCTGTCAGAGTTGCCGGGATCGCTTCCTAGAGCTCTT 1620

Qy       1681 CTACATGTATGATGAGGACGGCTATCAGTCCTACTGCACCGTGTGCTGTGAGGGCCGTGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTACATGTATGATGAGGACGGCTATCAGTCCTACTGCACCGTGTGCTGTGAGGGCCGTGA 1680

Qy       1741 ACTGCTGCTGTGCAGTAACACAAGCTGCTGCAGATGCTTCTGTGTGGAGTGTCTGGAGGT 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 ACTGCTGCTGTGCAGTAACACAAGCTGCTGCAGATGCTTCTGTGTGGAGTGTCTGGAGGT 1740

Qy       1801 GCTGGTGGGCGCAGGCACAGCTGAGGATGCCAAGCTGCAGGAACCCTGGAGCTGCTATAT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCTGGTGGGCGCAGGCACAGCTGAGGATGCCAAGCTGCAGGAACCCTGGAGCTGCTATAT 1800

Q        1861 GTGCCTCCCTCAGCGCTGCCATGGGGTCCTCCGACGCAGGAAAGATTGGAACATGCGCCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GTGCCTCCCTCAGCGCTGCCATGGGGTCCTCCGACGCAGGAAAGATTGGAACATGCGCCT 1860

Qy       1921 GCAAGACTTCTTCACTACTGATCCTGACCTGGAAGAATTTGAGCCACCCAAGTTGTACCC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCAAGACTTCTTCACTACTGATCCTGACCTGGAAGAATTTGAGCCACCCAAGTTGTACCC 1920

Qy       1981 AGCAATTCCTGCAGCCAAAAGGAGGCCCATTAGAGTCCTGTCTCTGTTTGATGGAATTGC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AGCAATTCCTGCAGCCAAAAGGAGGCCCATTAGAGTCCTGTCTCTGTTTGATGGAATTGC 1980

Qy       2041 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTGGAAAAGTACATTGCCTCCGA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTGGAAAAGTACATTGCCTCCGA 2040

Qy       2101 AGTCTGTGCAGAGTCCATCGCTGTGGGAACTGTTAAGCATGAAGGCCAGATCAAATATGT 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 AGTCTGTGCAGAGTCCATCGCTGTGGGAACTGTTAAGCATGAAGGCCAGATCAAATATGT 2100

Qy       2161 CAATGACGTCCGGAAAATCACCAAGAAAAATATTGAAGAGTGGGGCCCGTTCGACTTGGT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CAATGACGTCCGGAAAATCACCAAGAAAAATATTGAAGAGTGGGGCCCGTTCGACTTGGT 2160

Qy       2221 GATTGGTGGAAGCCCATGCAATGATCTCTCTAACGTCAATCCTGCCCGCAAAGGTTTATA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GATTGGTGGAAGCCCATGCAATGATCTCTCTAACGTCAATCCTGCCCGCAAAGGTTTATA 2220

         2281 TGAGGGCACAGGAAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCCGCCCCAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         2221 TGAGGGCACAGGAAGGCTCTTCTTCGAGTTTTACCACTTGCTGAATTATACCCGCCCCAA 2280

Qy       2341 GGAGGGCGACAACCGTCCATTCTTCTGGATGTTCGAGAATGTTGTGGCCATGAAAGTGAA 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GGAGGGCGACAACCGTCCATTCTTCTGGATGTTCGAGAATGTTGTGGCCATGAAAGTGAA 2340

Qy       2401 TGACAAGAAAGACATCTCAAGATTCCTGGCATGTAACCCAGTGATGATCGATGCCATCAA 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TGACAAGAAAGACATCTCAAGATTCCTGGCATGTAACCCAGTGATGATCGATGCCATCAA 2400

Qy       2461 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCTACCCGGAATGAACAGGCC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GGTGTCTGCTGCTCACAGGGCCCGGTACTTCTGGGGTAACCTACCCGGAATGAACAGGCC 2460

Qy       2521 CGTGATGGCTTCAAAGAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTCAGTAGGAC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CGTGATGGCTTCAAAGAATGATAAGCTCGAGCTGCAGGACTGCCTGGAGTTCAGTAGGAC 2520

Qy       2581 AGCAAAGTTAAAGAAAGTGCAGACAATAACCACCAAGTCGAACTCCATCAGACAGGGCAA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 AGCAAAGTTAAAGAAAGTGCAGACAATAACCACCAAGTCGAACTCCATCAGACAGGGCAA 2580

Qy       2641 AAACCAGCTTTTCCCTGTAGTCATGAATGGCAAGGACGACGTTTTGTGGTGCACTGAGCT 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AAACCAGCTTTTCCCTGTAGTCATGAATGGCAAGGACGACGTTTTGTGGTGCACTGAGCT 2640

         2701 CGAAAGGATCTTCGGCTTCCCTGCTCACTACACGGACGTGTCCAACATGGCCGCGGCGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         2641 CGAAAGGATCTTCGGCTTCCCTGCTCACTACACGGACGTGTCCAACATGGCCGCGGCGC 2700

Qy       2761 CCGTCAGAAGCTGCTGGGCAGGTCCTGGAGTGTACCGGTCATCAGACACCTGTTTGCCCC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 CCGTCAGAAGCTGCTGGGCAGGTCCTGGAGTGTACCGGTCATCAGACACCTGTTTGCCCC 2760

Qy       2821 CTTGAAGGACTACTTTGCCTGTGAATAGTTCTACCCAGGACTGGGGAGCTCTCGGTCAGA 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 CTTGAAGGACTACTTTGCCTGTGAATAGTTCTACCCAGGACTGGGGAGCTCTCGGTCAGA 2820

Qy       2881 GCCAGTGCCCAGAGTCACCCCTCCCTGAAGGCACCTCACCTGTCCCCTTTTTAGCTCACC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 GCCAGTGCCCAGAGTCACCCCTCCCTGAAGGCACCTCACCTGTCCCCTTTTTAGCTCACC 2880

Qy       2941 TGTGTGGGGCCTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGAGCAGAGCCTC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 TGTGTGGGGCCTCACATCACTGTACCTCAGCTTTCTCCTGCTCAGTGGGAGCAGAGCCTC 2940

Qy       3001 CTGGCCCTTGCAGGGGAGCCCCGGTGCTCCCTCCGTGTGCACAGCTCAGACCTGGCTGCT 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CTGGCCCTTGCAGGGGAGCCCCGGTGCTCCCTCCGTGTGCACAGCTCAGACCTGGCTGCT 3000

Qy       3061 TAGAGTAGCCCGGCATGGTGCTCATGTTCTCTTACCCTGAAACTTTAAAACTTGAAGTAG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TAGAGTAGCCCGGCATGGTGCTCATGTTCTCTTACCCTGAAACTTTAAAACTTGAAGTAG 3060

Qy       3121 GTAGTAAGATGGCTTTCTTTTACCCTCCTGAGTTTATCACTCAGAAGTGATGGCTAAGAT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GTAGTAAGATGGCTTTCTTTTACCCTCCTGAGTTTATCACTCAGAAGTGATGGCTAAGAT 3120

Qy       3181 ACCAAAAAAACAAACAAAAACAGAAACAAAAAACAAAAAAAAACCTCAACAGCTCTCTTA 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 ACCAAAAAAACAAACAAAAACAGAAACAAAAAACAAAAAAAAACCTCAACAGCTCTCTTA 3180

Qy       3241 GTACTCAGGTTCATGCTGCAAAATCACTTGAGATTTTGTTTTTAAGTAACCCGTGCTCCA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 GTACTCAGGTTCATGCTGCAAAATCACTTGAGATTTTGTTTTTAAGTAACCCGTGCTCCA 3240

Qy       3301 CATTTGCTGGAGGATGCTATTGTGAATGTGGGCTCAGATGAGCAAGGTCAAGGGGCCAAA 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CATTTGCTGGAGGATGCTATTGTGAATGTGGGCTCAGATGAGCAAGGTCAAGGGGCCAAA 3300

Qy       3361 AAAAATTCCCCCTCTCCCCCCAGGAGTATTTGAAGATGATGTTTATGGTTTAAGTCTTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 AAAAATTCCCCCTCTCCCCCCAGGAGTATTTGAAGATGATGTTTATGGTTTAAGTCTTCC 3360

Qy       3421 TGGCACCTTCCCCTTGCTTTGGTACAAGGGCTGAAGTCCTGTTGGTCTTGTAGCATTTCC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 TGGCACCTTCCCCTTGCTTTGGTACAAGGGCTGAAGTCCTGTTGGTCTTGTAGCATTTCC 3420

Qy       3481 CAGGATGATGATGTCAGCAGGGATGACATCACCACCTTTAGGGCTTTTCCCTGGCAGGGG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 CAGGATGATGATGTCAGCAGGGATGACATCACCACCTTTAGGGCTTTTCCCTGGCAGGGG 3480

Qy       3541 CCCATGTGGCTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAAGGGTGGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 CCCATGTGGCTAGTCCTCACGAAGACTGGAGTAGAATGTTTGGAGCTCAGGAAGGGTGGG 3540

Qy       3601 TGGAGTGGCCCTCTTCCAGGTGTGAGGGATACGAAGGAGGAAGCTTAGGGAAATCCATTC 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 TGGAGTGGCCCTCTTCCAGGTGTGAGGGATACGAAGGAGGAAGCTTAGGGAAATCCATTC 3600

Qy       3661 CCCACTCCCTCTTGCCAAATGAGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCCACTCCCTCTTGCCAAATGAGGGGCCCAGTCCCCAACAGCTCAGGTCCCCAGAACCCC 3660

Qy       3721 CTAGTTCCTCATGAGAAGCTAGGACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 CTAGTTCCTCATGAGAAGCTAGGACCAGAAGCACATCGTTCCCCTTATCTGAGCAGTGTT 3720

Qy       3781 TGGGGAACTACAGTGAAAACCTTCTGGAGATGTTAAAAGCTTTTACCCCACGATAGATT 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TGGGGAACTACAGTGAAAACCTTCTGGAGATGTTAAAAGCTTTTACCCCACGATAGATT 3780
```

```
Qy  3841 GTGTTTTTTAAGGGGGTGCCTTTTTTTTTTTAGGGGCATCACTGGAGATAAGAAAGCTGCATTTCAG 3900
         |||||||||||||||| || |||||||||| |||||||||||||||||||||||||||||||||||
Db  3781 GTGTTTTTTAAGGGGTGCTTTTTTTTTTTTTAGGGGCATCACTGGAGATAAGAAAGCTGCATTTCAG 3840

Qy  3901 AAATGCCATCGTAATGGTTTTTTAAACACCTTTTACCTAATTACAGGGTGCTATTTTATATAGA 3960
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841 AAATGCCATCGTAATTGGTTTTTTAAACACCTTTTACCTAATTACAGGGTGCTATTTTATATAGA 3900

Qy  3961 AGCAGACAACACTTCTTTTTTTATGACTCTCAGACTTTCTATTTTCATGTTACCATTTTTTTTT 4020
         |||||||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||
Db  3901 AGCAGACAACACTTCTTTTTTTAATGACTCTCAGACTTCTATTTTCATGTTACCATTTTTTTT 3960

Qy  4021 GTAACTCGCAAGGGTGTGGGCTTTTGTAAACTTCACAGGGTGTGGGGAGGAGGACTGCCTTGTTT 4080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961 GTAACTCGCAAGGGTGTGGGGCTTTTGTAAACTTCACAGGGTGTGGGGAGGAGGACTGCCTTGTTT 4020

Qy  4081 CAACAGTTTGTCTCCCACTGGTTTCTAATTTTTTAGGTGCAAAAGATGACAGATGCCCAGAGT 4140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021 CAACAGTTTGTCTCCCACTGGTTTCTAATTTTTTAGGTGGCAAAAGATGACAGATGCCCAGAGT 4080

Qy  4141 TTACCTTTCTGGTTGGATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAAA 4195
         |||||||||||||||||||||| ||||||||||||||||||||||||||
Db  4081 TTACCTTTCTGGTTGCATTAAAGTTGTATTTCTCTAAAAAAAAAAAAAAAAAAAAAA 4135
```

```
OM nucleic - nucleic search, using sw model

Run on:   July 16, 2003, 22:42:04 ; Search time 7176.89 Seconds
                                     (without alignments)
                                     17408.415 Million cell updates/sec

          GenCore version 5.1.6
          Copyright (c) 1993 - 2003 Compugen Ltd.

Title:          us-09-720-086-3
Perfect score:  4293
Sequence:       1 gccgcggcaccaggcgcgc.........taaaaaaaaaaaaaaaa 4293

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2054640 seqs, 1455140287B residues          4109280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  GenEmbl:*
            1:  gb_ba:*
            2:  gb_htg:*
            3:  gb_in:*
            4:  gb_om:*
            5:  gb_ov:*
            6:  gb_pat:*
            7:  gb_ph:*
            8:  gb_pl:*
            9:  gb_pr:*
           10:  gb_ro:*
           11:  gb_sts:*
           12:  gb_sy:*
           13:  gb_un:*
           14:  gb_vi:*
           15:  em_ba:*
           16:  em_fun:*
           17:  em_hum:*
           18:  em_in:*
           19:  em_mu:*
           20:  em_om:*
           21:  em_or:*
           22:  em_ov:*
           23:  em_pat:*
           24:  em_ph:*
           25:  em_pl:*
           26:  em_ro:*
           27:  em_sts:*
           28:  em_un:*
           29:  em_vi:*
           30:  em_vl:*
           31:  em_htg_hum:*
           32:  em_htg_inv:*
           33:  em_htg_other:*
           34:  em_htg_mus:*
           35:  em_htg_pln:*
           36:  em_htg_rod:*
           37:  em_htg_mam:*
           38:  em_htg_vrt:*
           39:  em_sy:*
           40:  em_htgo_hum:*
           41:  em_htgo_mus:*
           41:  em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4180.6	97.4	4258	9	AF331856	AF331856 Homo sapi
2	3005	70.0	3005	9	AF067972	AF067972 Homo sapi
3	2879.2	67.1	4192	10	AF068625	AF068625 Mus muscu
4	2825.4	65.8	4094	10	BC007466	BC007466 Mus muscu
5	2139.6	49.8	2191	9	AK025230	AK025230 Homo sapi
6	1667	38.8	1758	9	BC018214	BC018214 Homo sapi
7	1408.6	32.8	197652	9	AC012074	AC012074 Homo sapi
8	1221.4	28.5	123936	2	AC120824	AC120824 Rattus no
9	1157.6	27.0	168651	9	AC009474	AC009474 Homo sapi
c 10	1078.4	25.1	176697	2	AC112586	AC112586 Rattus no
11	791.8	18.4	4267	9	AF176228	AF176228 Homo sapi
12	788.2	18.4	4135	10	AF068627	AF068627 Mus muscu
13	788.2	18.4	4163	10	AF151974	AF151974 Mus muscu
14	786.6	18.3	4278	10	AF151970	AF151970 Mus muscu
15	764.4	17.8	4145	9	AF156488	AF156488 Homo sapi
16	764.4	17.8	4335	9	AF331857	AF331857 Homo sapi
17	762	17.7	4195	10	AF068626	AF068626 Mus muscu
18	760.4	17.7	4223	10	AF151973	AF151973 Mus muscu
19	760.4	17.7	4338	10	AF151969	AF151969 Mus muscu
c 20	694	16.2	119630	2	AC120040	AC120040 Rattus no
c 21	671.6	15.6	117264	2	AC103454	AC103454 Rattus no
c 22	647	15.1	204534	2	AC118195	AC118195 Mus muscu
23	647	15.1	208874	2	AC111092	AC111092 Mus muscu
24	601.4	14.0	3897	9	AF156487	AF156487 Homo sapi
25	600	14.0	3946	10	AF068628	AF068628 Mus muscu
26	600	14.0	3974	10	AF151976	AF151976 Mus muscu
27	598.4	13.9	4089	10	AF151971	AF151971 Mus muscu
28	573.8	13.4	4006	10	AY078427	AY078427 Mus muscu
29	572.2	13.3	4034	10	AF151975	AF151975 Mus muscu
30	572.2	13.3	4149	10	AF151972	AF151972 Mus muscu
31	544	12.7	2848	5	AF135438	AF135438 Danio rer
c 32	533.2	12.4	210269	2	AC116459	AC116459 Mus muscu
33	492.6	11.5	119630	2	AC112040	AC112040 Rattus no
34	492.6	11.5	3017	10	AK001191	AK001191 Homo sapi
35	417	9.7	167568	2	AC111734	AC111734 Rattus no
36	411.2	9.6	2127	9	AF129267	AF129267 Homo sapi
37	324.6	7.6	2077	6	AR129189	AR129189 Sequence
38	294.8	6.9	2057	9	AF129268	AF129268 Homo sapi
c 39	273	6.4	273	6	AX332056	AX332056 Sequence
c 40	273	6.4	273	6	AX411240	AX411240 Sequence
c 41	273	6.4	273	11	G06200	G06200 human STS w
c 42	254.8	5.9	56798	2	AC034293	AC034293 Homo sapi
c 43	252	5.9	2008	9	AF129269	AF129269 Homo sapi
c 44	190.6	4.4	204534	2	AC118195	AC118195 Mus muscu
c 45	190.6	4.4	208874	2	AC111092	AC111092 Mus muscu

ALIGNMENTS

```
RESULT 1
AF331856
LOCUS       AF331856                4258 bp    mRNA    linear   PRI 02-JAN-2002
DEFINITION  Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,
            complete cds.
ACCESSION   AF331856
VERSION     AF331856.1  GI:18033252
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4258)
  AUTHORS   N,J.; Pradhan,S. and Roberts,R.J.
  TITLE     Cloning, expression and characterization of human DNMT3 genes
```

Pred. No. is the number of results predicted by chance to have a

```
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 4258)
  AUTHORS  Ni,J., Pradhan,S. and Roberts,R.J.
  TITLE    Direct Submission
  JOURNAL  Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
           Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1..4258
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     gene            1..4258
                     /gene="DNMT3A"
     CDS             227..2956
                     /gene="DNMT3A"
                     /codon_start=1
                     /product="DNA cytosine methyltransferase 3 alpha"
                     /protein_id="AAL57039.1"
                     /db_xref="GI:18033253"
                     /translation="MPSSGPGDTSSSAAEREDRKDGEEQEEPRGKEERQEPSTTARK
                     VGRPGRRKRHPPVESGDTPKDPAVISKSPSMAQDSGASELLPNGDLEKRSEPQPEGS
                     PAGGCKGRAPAEEGEAAETLPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIESMKM
                     EGSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISRRKRDEMLARWKREAEKKAKVI
                     AGMNAVEENGPGESQKVEEASPPAVQQPTDPASPTVATTPEPVGSDAGDKNATKAGD
                     DEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMMTGRSRAAEEGTRWVMWFGDGKFS
                     VVCVEKLMPLSSFSSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLPPVCHDSDESDT
                     AKAVEVQNKPMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPEAAAYAPPPP
                     AKKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVTLEHPLFVG
                     GMCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFCVECVDLLVGP
                     GAAQAAIKEDPWNCYMCGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYPPVP
                     AEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVG
                     DVRSVTQKHIQEWGPFDLVIGGSPCNDLSIIVNPARKGLYEGTGRLFFEYYRLLHDARP
                     KEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPCM
                     NRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVFMNEKEDIL
                     WCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSMSVPVIRHLFAPLKEYFACV"
BASE COUNT     1062 a   1089 c   1257 g    850 t
ORIGIN
```

```
Query Match
 Best Local Similarity    97.4%;
 Matches 4210;  Conservative  0;  Mismatches  0;

 Similarity    99.7%;  Pred. NO. 0;
 Score 4180.6;   DB 9;   Length 4258;
 Indels  4;   Gaps  2;
```

```
Qy    52 GGTGGGAGCCATCGAAGCCCCCCCACCACCAGCGCTGACAGAGGCACACGTTCACCAGAGGCGCTC  111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    40 GGTGGGAGCCATCGAAGCCCCCCCACCACCAGCGCTGACAGAGGCACACGTTCACCAGAGGCGCTC   99

Qy   112 AACACCCGGGATCTATGTTTAAGTTTTTAAACTCTCGGCCTCCAAAAGACCACGGATAATTCCTTC  171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   100 AACACCCGGGATCTATGTTTAAGTTTTTAAACTCTCGGCCTCCAAAAGACCACGGATAATTCCTTC  159

Qy   172 CCCAAAAGCCCAGCAGCACCCCCCCCCCAGCCCCCGGCCCAGCCCCAGCCCGGCGCGCCCAGAT   231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   160 CCCAAAAGCCCAGCAGCACCCCCCCCCCAGCCCCCGGCCCAGCCCCAGCCCGGCGCGCCCAGAT   219

Qy   232 GCCCGCCATGCCCTCCAGCGCCGGCCCCCGGGGACACACCAGCAGCTCTGCTTGCCGAGCGGGAGGA  291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   220 GCCCGCCATGCCCTCCAGCGCCGGCCCCCGGGGACACACCAGCAGCTCTGCTTGCCGAGCGGGAGGA  279

Qy   292 GGACCGAAAGGACGGAGAGGAGCAGGAGGAGCCGCGTGGCCAAGGACGCGCCAAGAGACC  351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   280 GGACCGAAAGGACGGAGAGGAGCAGGAGGAGCCGCGTGGCCAAGGACGCGCCAAGAGACC  339

Qy   352 CAGCACCACCGGCCACGGAAGGTGGGGCGGGGGGGCCTGGGAGGAGAAGCGAAGCACCCCCCCGTCGA  411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   340 CAGCACCACCGGCCACGGAAGGTGGGGCGGGGGGGCCTGGGAGGAGAAGCGAAGCACCCCCCCGTGGA  399

Qy   412 AAGCCGGTGACACGCCAAAAGGACCCCTGCGGGTGGATCTCCAAGTCCCCATCCCATGGCCCAGGA  471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   400 AAGCCGGTGACACGCCAAAAGGACCCCTGCGGGTGGATCTCCAAGTCCCCATCCCATGGCCCAGGA  459

Qy   472 CTCAGCGCGGCCTCGAGAGCTATTACCCCAATGGGGACTTGGAGAAACGGGAGTGAGCCCCAGCC  531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   460 CTCAGGGCGGCCTCAGAGCTATTACCCCAATGGGGACTTGGAGAAACGGGAGTGAGCCCCAGCC  519
```

```
Qy   532 AGAGGAGGGGGAGGGGGAGCCCCCTGCTTGGGGGGCGGGGGGGCCCCCAGCAGAGGGAGAGGGAGGTGC  591
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   520 AGAGGAGGGGGAGGGGGAGCCCCCTGCTTGGGGGGCGGGGGGGCCCCCAGCAGAGGGAGAGGGAGGTGC  579

Qy   592 AGCTGCAGACCCTGCCTGAAGCCTCAGAGCACTGTGGAGAAAAATGGCTGCTGCACCCCCAAGGA  651
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   580 AGCTGAGAGACCCTGCCTGAAGCCTCAGAGCACTGTGGAGAAAAATGGCTGCTGCACCCCCAAGGA  639

Qy   652 GGGGCCGAGGAGCCCCCTGCAGAGAACGGGCCAAAGAAACAGAAGGGAGACCAACATGGAATCCAT  711
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   640 GGGGCCGAGGAGCCCCCTGCAGAGAACGGGCCAAAGAAACAGAAGGGAGACCAACATCGAATCCAT  699

Qy   712 GAAAAATGGAAGGGGCCTCCCCGGGGGGGCCCGGCTTGGGGGGCTTGGGGCTGGGAGTCCAGCCTCCG  771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   700 GAAAAATGGAAGGGGCCTCCCCGGGGGGCCCGGCTTGGGGGGCTTGGGGCTGGGAGTCCAGCCTCCG  759

Qy   772 TCAGCGCGCCCACGCCCGAGGCCTCACCCTTCCCAGCGCGCGGGGGGGGACCCCCCTACTACAATCAGCAAGCG  831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   760 TCAGCGCGCCCATGCCCGAGGCTCACCTTCCCAGCGCGCGGGGGGGGACCCCCTACTACAATCAGCAAGCG  819

Qy   832 CAAGCGGGACGAGTGGCTGCACGCGCTGGAAGGAGGCTGAGAAGAAAAGCCAAGGTCAT  891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   820 CAAGCGGGACGAGTGGCTGCACGCGCTGGAAGGAGGCTGAGAAGAAAAGCCAAGGTCAT  879

Qy   892 TGCAGGAATGAATGCTCGTGGAAGAAAAACCAGGGGCCCGGGGGAGTCTCAGAAGGTGGAGGA  951
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   880 TGCAGGAATGAATGCTCGTGGAAGAAAAACCAGGGGGCCCGGGGGAGTCTCAGAAGGTGGAGGA  939

Qy   952 GGCCCAGCCCCCTCCTCGTCGTCCAGCACCCCACTCGCCCCGCCATCCCCCCACTCTGGGCTTACCAC  1011
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   940 GGCCAGCCCCTCCTCGTCGTCCAGCACCCACTGCCCCGCCATCCCCCACTGTGGGCTTACCAC  999

Qy  1012 GGCCGAGCCCGTGGGGTCCGGATGCTGGGACACAAGAATGCCACCAAAGCAGGCGGATGACGA  1071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1000 GGCTGAGCCCGTGGGGTCCGGATGCTGGGACACAAGAATGCCACCAAAGCAGGCGGATGACGA  1059

Qy  1072 GCCAGAGTACGAGGACGGCCCCCGGGGCTTTGGCATTGGGGAGCTGGTGTGGGGAAACTGCG  1131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1060 GCCAGAGTACGAGGACGGCCCCCGGGGCTTTGGCATTGGGGAGCTGGTGTGGGGAAACTGCG  1119

Qy  1132 GGGCTTCTCCTGGTGGCCAGGGCGCCATTGTGTCTTGGTGGATGATGACGGGCCGGAGCCGAGC  1191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1120 GGGCTTCTCCTGGTGGCCAGGGCGCCATTGTGTCTTGGTGGATGATGACGGGCCGGAGCCGAGC  1179

Qy  1192 AGCTGAAGGCACCCGCTGGGTCATGTGGTTCGGAGACGGCAAAATTCTCAGTGGTGTGTGT  1251
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1180 AGCTGAAGGCACCCGCTGGGTCATGTGGTTCGGAGACGGCAAAATTCTCAGTGGTGTGTGT  1239

Qy  1252 TGAGAAGCTGATGCCGCTGAGCTCGTTTTGCAGTGCGTTCCACCAGGCCACAGTTACAAACA  1311
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1240 TGAGAAGCTGATGCCGCTGAGCTCGTTTTGCAGTGCGTTCCACCAGGCCACAGTTACAAACA  1299

Qy  1312 GCAGCCCATGTACCGCAAAAGCCATCTACGCAGCCTCTGCCAGGTGGCCGCAGAGCCGGTGG  1371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1300 GCAGCCCATGTACCGCAAAAGCCATCTACGCAGCCTCTGCCAGGTGGCCGCAGAGCCGGGGG  1359

Qy  1372 GAAGCTGTTCCCGGGTGTGCCACGACACGATAGAGTGACACTGCCAAGGCCGTGGGAGGT  1431
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1360 GAAGCTGTTCCCGGGTGTGCCACGACACGATAGAGTGACACTGCCAAGGCCGTGGGAGGT  1419

Qy  1432 GCGAGAACAAGCCCCATGATTGAATGGGCCCTGGGGGGCCTTCCAGCCTTTCTGGCCCTAAGGG  1491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1420 GCGAGAACAAGCCCCATGATTGAATGGGCCCTGGGGGGCCTTCCAGCCTTTCTGGCCCTAAGGG  1479

Qy  1492 CCTGGAGCCACCCAGAGAGAAGAAGAATCCCTACAAAAGAAAGTGTACACGGACATGTGGGT  1551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1480 CCTGGAGCCACCCAGAGAGAAGAAGAATCCCTACAAAAGAAAGTGTACACGGACATGTGGGT  1539

Qy  1552 GGAACCTGAGGCAGCTCCCTACGCACCACCTCCACCACGGCCAAAAAAGCCCCGGGAAGAGCAC  1611
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1540 GGAACCTGAGGCAGCTCCCTACGCACCACCTCCACCACGGCCAAAAAAGCCCCGGGAAGAGCAC  1599
```

Qy   1612 AGCGGAGAAGCCCAAGGTTCAAGGAGAGAGATTATTTGATGAGCGACGGCCACAGAGAGAGCGGCTGGTGTA 1671
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1600 AGCGGAGAAGCCCAAGGTTCAAGGAGAGAGATTATTTGATGAGCGACGGCCACAGAGAGAGCGGCTGGTGTA 1659

Qy   1672 CGAGGTGCGGCCGGAAGAAGTGCCGGAACATTGAGACGAGCCTCCTTGTGGGGAGCCTCAA 1731
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1660 CGAGGTGCGGCCGGAAGAAGTGCCGGAACATTGAGACGAGCCTCCTTGTGGGGAGCCTCAA 1719

Qy   1732 TGTTTACCCTGGAACACCCCCTCTTCGTTGGGAGGAATGTGGCCAAAACTGCAAGAAACTGCTT 1791
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1720 TGTTTACCCTGGAACACCCCCTCTTCGTTGGGAGGAATGTGGCCAAAACTGCAAGAAACTGCTT 1779

Qy   1792 TCTGGAGTGTGCGGTACCAGTACGACGACGACGGCTACCAGTACGACTGCACCATCTGGTG 1851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1780 TCTGGAGTGTGCGGTACCAGTACGACGACGACGGCTACCAGTACGACTGCACCATCTGGTG 1839

Qy   1852 TGGGGGGCCGTGAGGTGCTCATGTGCGGAAAACACAACTGCTGCAGGTGCTTTTGCGTGGA 1911
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1840 TGGGGGGCCGTGAGGTGCTCATGTGCGGAAAACACAACTGCTGCAGGTGCTTTTGCGTGGA 1899

Qy   1912 GTGTGTGGACCTCTTGGTGGGGCCGGGGGGCCGGGGGGCCCAGGGCAGCCATTAAGGAGACCCCTG 1971
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1900 GTGTGTGGACCTCTTGGTGGGGCCGGGGGGCCGGGGGGCCCAGGGCAGCCATTAAGGAGACCCCTG 1959

Qy   1972 GAACTGCTACATGTGCGGGGCACAAGGGGTACCTACGGGGCTGCTGCGGGGCGAGAGGACTG 2031
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1960 GAACTGCTACATGTGCGGGGCACAAGGGGTACCTACGGGGCTGCTGCGGGGCGAGAGGACTG 2019

Qy   2032 GCCCTCCCGGCTCCAGATGTTCTTCGCTAATAACCACGACGCAGGAATTTGACCCTCAAA 2091
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2020 GCCCTCCCGGGCTTCCAGATGTTCTTCGCTAATAACCACGACCAGGAATTTGACCCTCAAA 2079

Qy   2092 GGTTTACCCACCTGTCCCAGCTGTCGAGAAGAGGAAGCCCATCCGGGGTCGTGTCTCTCTTTGA 2151
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2080 GGTTTACCCACCTGTCCCAGCTGTCGAGAAGAGGAAGCCCATCCGGGGTCGTGTCTCTCTTTGA 2139

Qy   2152 TGGGAATCGCTACAGCGGCTCCCTGGTGCTCAAAGGACTTGGGGACTTGGGACGGCCGCTACAT 2211
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2140 TGGGAATCGCTACAGGGGCTCCCTGGTGCTGAAGGACTTGGGCATGGTGGACGGCCGCTACAT 2199

Qy   2212 TGCCTCGGGAGGTGGTGTCGAGGGAGCTCCATCACGTGGGGACCCAGGGGAAGAT 2271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2200 TGCCTCGGGAGGTGTGTGTGAGGGAGCTCCATCACGGTGGGGCACCAGGGGAAGAT 2259

Qy   2272 CATGTACGTCGGGAGCCGTCCGCACAGAGAGCCAGCGTCACACAGAGTGGGGGCCCATT 2331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2260 CATGTACGTCGGGAGCCGTCCGGAGCCGTCACACACAGCAATATCCAGAGTGGGGGCCCATT 2319

Qy   2332 CGAGCTGGTTGATTGGGGGACGTCCGGCAGTCCCTGCGAAATGACCTCTCCATCCGTCAAACCTGCTCGCAA 2391
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2320 CGATCTGGTTGATTGGGGGACGTCCGGCAGTCCCTGCGAAATGACCTCTCCATCCGTCAAACCCTGCTCGCAA 2379

Qy   2392 GGGGCCTCTTACGAGGGGCACTGGGCCGCGGGGCCCGCCCGCCCTACTTCTCTTTGCGCATGATGC 2451
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2380 GGGGCCTTCTACGAGGGGCACTGGGCCGCGGGGCCCGCCCGCCTCTTCTTCGAGTTCTTACCCATGATGC 2439

Qy   2452 GCCGGGCCCAAGGAGGAGGGAGGATGGAGGATGATCGCCCCCCTTCTTCCTTCTTGGGCTCTTTGAGAAGTGTGGCCCAT 2511
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2440 GCCGGGCCCAAGGAGGAGGGAGGATGAATCGCCCCCTTCTCTTCTGGCTCTTTTGAGAATGTGGGCCCAT 2499

Qy   2512 GGGCGTTAGTGACAAGAGGGACATCTCCGGATTTCTCGAGTGTCCAACCCTGTGATGATTGA 2571
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2500 GGGCGTTAGTGACAAGAGGGACATCTCCGGATTTCTCGAGTGTCCAACCCTGTGATGATTGA 2559

Qy   2572 TGCCAAAGAAGTGTCAGCTGCACACAGGGCCCGCCCTACTTCTCTGGGGTAAGCTTCCCGGTAT 2631
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2560 TGCCAAAGAAGTTCAGCTGCACACAGGGCCCGCCCTACTTCTCTGGGGTAAGCTTCCCGGTAT 2619

Qy   2632 GAACAGGCCCGTTGGCATCCACCGTGTGAAGTGATGAATGAAGATAACCTCAGGAGTGTCTGGAGCA 2691
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2620 GAACAGGCCGTTGGCATCCACTCTGTGAAATGATAAGCTGGACTGCAGCAGAGTGTCTGGACCA 2679

Qy   2692 TGGCAGGAGATAGCCAAGTTCAGTGAAAAGTGAGGACCATTACTACGAGGTCAAACTCCATAAA 2751
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2680 TGGCAGGAGATAGCCAAGTTCAGTGAAAAGTGAGGACCATTACTACGAGGTCAAACTCCATAAA 2739

Db   2680 TGGCAGGAGATAGCCAAGTTCAGTGAAAAGTGAGGACCATTACTACGAGGTCAAACTCCATAAA 2739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2752 GCAGGGCAAAGACCAGCATTTTCCTCGTCTTCATGAATGAGAAGAGGACATCTTATGGTG 2811
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2740 GCAGGGCAAAGACCAGCATTTTCCTCTGTCTTCATGAATGAGAAGAGGACATCTTATGGTG 2799

Qy   2812 CACTGAAAATGGAAAAGGGTATTTTGGTTTCCCAGTCCCACTATACTGACGTCTCCAAACATGAG 2871
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2800 CACTGAAAATGGAAAAGGGTATTTTGGTTTCCCAGTCCCACTATACTGACGTCTCCAAACATGAG 2859

Qy   2872 CCGCTTGGGCGAGGCAGAGACTGCTGGGGCCGGGTCATGGAGCGGTGCCAAACTGCAACA 2931
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2860 CCGCTTGGGCGAGGCAGAGACTGCTGGGGCCGGGTCATGGAGCGGTGCCAAACTGCAACA 2919

Qy   2932 AGAGGCGGACAAAGTTAAACAAACAAAAACAAAAAACATAATAAAAAACACCAAGAACA 2991
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2920 CTTCGCTCCGCTGAAGGAGTATTTTGCGTGTGTTGGAAGCAGCATGGGGGCAAACTGAGGT 2979

Qy   2992 AGCGGACACAAAGTTAAACAAACAAAAAACAAAAAACATAATAAAAAACACCAAGAACA 3051
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2980 AGCGGACACAAAGTTAAACAAACAAAAAACAAAAAACATAATAAAAAACACCAAGAACA 3039

Qy   3052 TGAGGAATGGGAGAAGTATTCACGCACCCAGAGAGAGAAAAAGGAATTTAAAAAACCAC 3111
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3040 TGAGGAATGGGAGAAGTATTCAGTGCACCCAGAGAGAGAAAAAGGAATTTAAAAAACCAC 3099

Qy   3112 AGAGGCGGAAATACCCGGAGGGGGCTTTGCCTTGCGAAAAAGGGTTGGAACATCATCTCCCGATT 3171
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3100 AGAGGCGGAAATACCGGAGGGGGCTTTGCCTTGCGAAAAAGGGTTGGACATCATCCCTGATT 3159

Qy   3172 TTTCAATGTTATTCTTCAGTCCTATTTAAAAACAAAAACCAAGCTCCCTTCCCTCCC 3231
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3160 TTTCAATGTTATTCTTCAGTCTCTATTTAAAAACAAAAAGCTCCCTTCCCTCCC 3219

Qy   3232 CCTTCCCTTTTTTTTTCGGTCAGACCTTTTTATTTTTACTCCTTTTTTCAGAGGGGGTTTTCTGT 3291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3220 CCTTCCCTTTTTTTTTCGGTCAGACCTTTTTATTTTTGCTGAAAACAAGAGTCCCCTTTTTCTGT 3279

Qy   3292 TTGTTTTGGGTTTTTCTTCTTGCTGTGTGGCTCAGGGGAAGGGTTATTGCAGGAAAATCAG 3351
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3280 TTGTTTTGGGTTTTTCTTCTTGCGTGTGACTGAAACAAGAGGTTAATTGCAGCAAAATCAG 3339

Qy   3352 TAACAAAAAAATAGTAACAATACCTTGCAGAGGAAAGGTGGGAGGAGGAAAAAGGGAA 3411
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3340 TAACAAAAAAATAGTAACAATACCTTGCAGAGGAAAGGTG---GGAGAGGGAAAAAAGGA 3396

Qy   3412 ATTTTTAAAGAAAATCTATAATATATATTGGGTTGTCTTTTTTGTTTTTTTTTTTTTT 3471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3397 AATTCTATAGAAAATCTATAATATATATTGGGTTTGTCTTTTTTTGTTTTTTTTTTT 3456

Qy   3472 GGGGTTTTTTTTTTTTTACTATATAATCTTTTTTTTTGTCTTTTTTTTTGATCAGGATGATAGGAG 3531
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3457 GGGGTTTTTT--TTTTTACTATATAATCTTTTTTTGTGTCTCTGGACGTTCTTGATCAGGATAGGAG 3515

Qy   3532 CACAAGCAGGGGACGGAAAGAGAGACAGAACTCAGCGGCCAATTCCCTCCCCAGCCACTG 3591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3516 CACAACCAGGGGACGGAAAGAGAGAGCACACTCAGCCGGCCACATTCCCTCCCCAGCCACTG 3575

Qy   3592 AGCTGTCGTGCCAGCACCATTCCTGGTCACGCCAAAACAGAACCCAGTTAGCAGCAGGGAG 3651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3576 AGCTGTCGTGCCAGCACCATTCCTGGTCACGCCAAAACAGAACCCAGTTAGCAGCAGGGAG 3635

Qy   3652 ACGAGAACACACACACAAGACACATTTTCTAACAGATTTTTTACAGGTGCCTTACCCACCACACAGGAAAC 3711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3636 ACGAGAACACACACACAAGACACATTTTCTACAGTTATTTCAGGTGCCTTACCACCACACAGGAAAC 3695

Qy   3712 CTTGAAGAAAATCAGTTTCTAGAAGCCGCGGCCGTGTTACCTCTTGTTTACGTTTATATAGTATATA 3771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3696 CTTGAAGAAAATCAGTTTCTAGAAGCCGCGGCCGTGTTACCTCTTGTTTATATATAGTATATATA 3755

Qy   3772 TGATAGATATATGAGATATATATATAAAAAGGTACTGTTAACTACTTGTTACAACCCGGACTTCAT 3831
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2

LOCUS   AF067972   3005 bp   mRNA   linear   PRI 12-FEB-2001
DEFINITION Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,
            complete cds.
ACCESSION   AF067972
VERSION   AF067972.2 GI:12746531
KEYWORDS   .
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3005)
  AUTHORS   Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.
            and Li,E.
  TITLE   Cloning, expression and chromosome locations of the human DNMT3
            gene family.
  JOURNAL   Gene 236 (1), 87-95 (1999)
   MEDLINE   99365304
   PUBMED   10433969
REFERENCE   2 (bases 1 to 3005)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th
            Street, Charlestown, MA 02129, USA
REFERENCE   3 (bases 1 to 3005)
  AUTHORS   Hata,K., Shirohzu,H., Sasaki,H. and En,L.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th
            Street, Charlestown, MA 02129, USA
REMARK   Sequence update by submitter
COMMENT   On Feb 12, 2001 this sequence version replaced gi:4927369.
FEATURES   Location/Qualifiers
     source   1..3005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p23; between D2S171 and D2S174; 48-50 cM"

     gene   1..3005
                     /gene="DNMT3A"
     CDS   230..2968
                     /gene="DNMT3A"
                     /function="de novo DNA cytosine methyltransferase"
                     /codon_start=1
                     /product="DNA cytosine methyltransferase 3 alpha"
                     /protein_id="AAD33084.2"
                     /db_xref="GI:12746532"
                     /translation="MPAMPSSGPGDTSSSAAEREEDRKDGEEQEEPRGKEEQEQPSTT
                     ARKVGRPGRKRKHPPVESGDTPKDPAVISKSPSMAQDSGASELLPNGDLEKRSEPQPE
                     EGSPAGGQKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIES
                     MKMEGSRGRLRGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKA
                     KVIAGMNAVEENQGPGESQKVEEASPPAVQQPTDPASPTVATTPEPVGSDAGDKNATK
                     AGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDG
                     KFSVVCVEKLMPLSSFCSEAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFPVCHDSDE
                     SDTAAVEVONKPMIEWALGGFQPSGPKGLEPPEEEKNPYKEVYTDMWVEPEAAAYAP
                     PPPAKKPRKSTAEKPKVKEIIDERTRRLVYEVRQKCRNIEDICISCGSLNVTLEHPL
                     FVGGMCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFVECVDLL
                     VGPAAQAAIKEDPWNCYMCGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYP
                     PVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM
                     YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEYYRLLHD
                     ARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL
                     PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSITKQGKDQHFPVFMNEKE
                     DILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV"

BASE COUNT   674 a   846 c   978 g   507 t
ORIGIN

Query Match   70.0%; Score 3005; DB 9; Length 3005;
Best Local Similarity   100.0%;
Matches 3005; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GCCGGCGGCCACCAGGGCGGGGCGGGCCGAGCCCGGCGGCCGGGGCCGACCCCACCCGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCCGGCGGCCACCAGGGCGGGGCGGGCCGAGCCCGGCGGCCGGGGCCGACCCCACCCGGG 60

Qy   61 ATCGAAGCCCCCACCCACCCAGGCTGACGAGAGCCACCGTTCACCAGAGGCGGCTCAACACCCGGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATCGAAGCCCCCACCCACCCAGGCTGACGAGAGCCACCGTTCACCAGAGGCGGCTCAACACCCGGG 120

Qy   121 ATCTATGTTTAAGTTTTAACTCTCGCCTCCAAAGACCACGATAATTCCTTCCCCAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATCTATGTTTAAGTTTTAACTCTCGCCTCCAAAGACCACGATAATTCCTTCCCCAAAGCC 180

Qy   181 CAGCAGCCCCCCAGCCCGGCCCAGCCCCTGCCTCCCGGGCCCCCCAGATGCCCGCCCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGCAGCCCCCCAGCCCGGCCCAGCCCCTGCCTCCCGGGCCCCCCAGATGCCCGCCCAT 240

Qy   241 GCCCTCCAGCGCGCCCCGGGAGACCAGCAGCCAGCAGCCAGCTCTCTGCCTGGAGGGGAGGGACCGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCCCTCCAGCGCGCCCCGGGAGACCAGCAGCCAGCAGCCAGCTCTCTGCCTGGAGGGGAGGGACCGAAA 300

Qy   301 GGACGGAGCAGGAGGACCAGGAGGAGGAGCCCGCGTTGGGCAAAGGAGCCCAGCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGACGGAGCAGGAGGACCAGGAGGAGGAGCCCGCGTTGGGCAAAGGAGCCCAGCCACCAC 360

Qy   361 GGCACGGAAAGGTGGGGCGGCCTGGGGAGGGAAACGCAAGCACCCCCCGGTGGAAAACGGGTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGCACGGAAAGGTGGGGCGGCCTGGGGAGGGAAACGCAAGCACCCCCCGGTGGAAAACGGGTGA 420

Qy   421 CACCCCAAAGGAGACCCTGCGGTGATCTCCAAGTCCCCATCCATTGGCCAGGACTCAGGCGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CACCCCAAAGGAGACCCTGCGGTGATCTCCAAGTCCCCATCCATTGGCCAGGACTCAGGCGC 480

Qy   481 CTCAGAGCTGTTACCCAATGGGGACACTTGGAGAAGCGGAGTGAGCCCCCCAGCCAGAGGAGGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CTCAGAGCTGTTACCCAATGGGGACACTTGGAGAAGCGGAGTGAGCCCCCCAGCCAGAGGAGGG 540

Qy   541 GAGCCCTGCCTGGGGGCAGAAGGGCGGGCCCCAGCAGAGGGAGAGGGTGCAGCTGAGAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAGCCCTGCCTGGGGGCAGAAGGGCGGGCCCCAGCAGAGGGAGAGGGTGCAGCTGAGAC 600

Qy   601 CCTGCCTGAAGCCTCAAGAGCAGTGGAAAATGGCCTGCTGCCACCCCCAAGAGAGGGCGAGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CCTGCCTGAAGCCTCAAGAGCAGTGGAAAATGGCCTGCTGCCACCCCCAAGAGAGGGCGAGG 660

Db 3756 TGATAGATATGAGATATATATATAAAAGGTACTGTTAACTACTGTACACACCGACTTCAT 3815
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3832 AATGGTGCTTTCAAACAGCGAGATCAGCGAGTAAAAACATCAGCTTCCCACGTTGCCTTCTCGCGC 3891

Db 3816 AATGGTGCTTTCAAACAGCGAGATCAGCGAGTAAAAACATCAGCTTCCCACGTTGCCTTCTCGCGC 3875
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3892 AAAGGGTTTCACCAAGGATGGAGAAAAGGGAGACAGCTTGCCAGTACGGCCGGTTCTCACGGT 3951

Db 3876 AAAGGGTTTCACCAAGGATGGAGAAAAGGGAGACAGCTTGCCAGTACGGCCGGTTCTCACGGT 3935
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 3952 GGGCTCTTCCCCTTGGTTTGTTAACGAAAGTGAAAGGAGGAGGAGAACTTGGGAGCCAGGTTCTCC 4011

Db 3936 GGGCTCTTCCCCTTGGTTTGTTAACGAAAGTGAAAGGAGGAGGAGAACTTGGGAGCCAGGTTCTCC 3995
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4012 CTGCCAAAAAGGGGGCTAGATGAGGTGGTGTCGGGGCCCGTGGGACAGCTGAGAGTGGGATTCA 4071

Db 3996 CTGCCAAAAAGGGGGCTAGATGAGGTGGTGTCGGGGCCCGTGGGACAGCTGAGAGTGGGATTCA 4055
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4072 TCCAGACTCATGCAATAAACCCTTTGATTGTTGTTTTCTAAAAGGAGAGACTCCCTCGGGCAAGATG 4131

Db 4056 TCCAGACTCATGCAATAAACCCTTTGATTGTTGTTTTCTAAAAGGAGAGACTCCCTCGGGCAAGATG 4115
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4132 GCAGAGGGGTACGGAGTCTTCAGGCCCCAGTTTCTCACTTTAGCCAATTCGAGGGCTGGAGCC 4191

Db 4116 GCAGAGGGGTACGGAGTCTTCAGGCCCCAGTTTCTCACTTTAGCCAATTCGAGGGCTGGAGCC 4175
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4192 TGGTGGGGATCAGAACTAATCCAGACTGTGGGAAAGTGACAGTCAAAAACCCCACCTGGAGCC 4251

Db 4176 TGGTGGGATCAGAACTAATCCAGACTGTGGGAAAGTGACAGTCAAAAACCCCACCTGGAGCC 4235
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4252 AAATAAAAAAAACATACAAAACGT 4274

Db 4236 AAATAAAAAAAACATACAAAACGT 4258

```
RESULT 3
LOCUS       AF068625                4192 bp    mRNA    linear   ROD 06-DEC-1999
DEFINITION  Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,
            complete cds.
ACCESSION   AF068625
VERSION     AF068625.2  GI:6449467
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4192)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE     Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
  JOURNAL   Nat. Genet. 19 (3), 219-220 (1998)
   MEDLINE   98324766
   PUBMED    9662389
REFERENCE   2  (bases 1 to 4192)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital,
            Charlestown, MA 02129, USA
REFERENCE   3  (bases 1 to 4192)
  AUTHORS   Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK      Sequence update by submitter
COMMENT     On Nov 18, 1999 this sequence version replaced gi:3327977.
FEATURES             Location/Qualifiers
     source          1..4192
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /map="4.0 cM"
     gene            1..4192
                     /gene="Dnmt3a"
     misc_feature    1..4192
                     /gene="Dnmt3a"
                     /note="similar to EST sequences deposited in GenBank
                     Accession Numbers AA052791, AA111043, AA154890, AA240794,
                     AA756653, W58898, W59229, W91664, and W91665"
     CDS             217..2943
                     /gene="Dnmt3a"
                     /function="de novo DNA methylation"
                     /note="contains Cys-rich region"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3A"
                     /protein_id="AAC40177.2"
                     /db_xref="GI:6449468"
                     /translation="MPSSGPGDTSSSSLEREDDRKEGEEQEENRKEEROEPSATARK
                     VRGPRKRKHPVESSDTPKDPAVTRKSQPMAQDSPSDLLPNGDLEKRSEPQPEEGS
                     PAAGQKGGAPAEGEGTEPPEASRAVENGCCVTKEGRGASAGEGKEQKQTNIESMKME
```

Sequence alignment section:

```
                    GSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKAKVIA
                    VMNAVEENQASGESQKVEEASPPAVQQPTDPASPTVATTPEPVGGDAGDKNATKAADD
                    EPEYEDGPGIGELVWGKLRGFSWWPGRIVSWWMPGRSRAAAEGTRWVMWFGDGKFSV
                    VCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFPACHDSSEDSDSG
                    KAVEVONKQMIEWALGGFQPSGPKGLEPEEEKNPYKEYVTDMWVEPEAAAVAPPPPA
                    KKPRKSTTEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVTLEHPLFIGG
                    MCQNCKNCFLECAYQYDDDGYQSYCTICCGGREVLMCGNNNCCRCFCVECCVDLLVGPG
                    AAQAAIKEDPWNCTWMCCHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYPPVPA
                    EKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIMYVGD
                    VRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEYYRLLHDARPK
                    EGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNLPGMN
                    RPLASTVNDKLELDECLEHGRIAKFSKVRTITTRSNSIKQGKDQHFPVPMNEKEDILW
                    CTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV
```

Alignment details (Query vs Db):

```
Query Match      67.1%;  Score 2879.2;  DB 10;  Length 4192;
Best Local Similarity   83.8%;
Matches 3579;  Conservative   0;  Mismatches   578;  Indels 113;  Gaps 24;

BASE COUNT      1094 a    1186 g     840 t
ORIGIN
```

```
Qy   24 CGGGCCCGGCCGCCGAACCCACCCGGCCCCATACGGTYGGAGCCCATCGAAAGCCCCACCCACACAGGCT 83
        |   | | |||||||||| || | | ||||||   |   |  || ||  |  | || | | |||
Db   17 CCGGGGCGGCCCGGAGCCCGGCGCGGGGCCCACACGGCAGAGGCGGCCCTGAAGCCCCAGGCTGAGGCT 76

Qy   84 GACAGAGAGGCACCCGTTCACCAGAGGGGCTCAACACACCGGGATCTPATGTTTAAGTTTTAACTCT 143
        | ||||||||||||| |||  |||||||||||||||||| |||| |  ||||||||||||   ||||
Db   77 G-------------CACTTTTCCGAGGGGCTTGACATCAGGGTCTTATGTTTAGTTTTAGCTCT 126

Qy  144 CGGCTTCCAAAAGACCACCGATAATTTCCTTTCCCCAAAAGCCC-AGCAGCCCCCAGC-CCCGCG 201
        | | | ||||||  ||| || |||| |||||||  ||||||| | | |  |||||||||||  ||
Db  127 TGCTTACAAAAGACCACGGCAATTCCTTCTCGTGAAAGCCCTCGGCAGCCCCACCAGGCCCCTCG 186

Qy  202 CAGCCCAGCCTGCCTCCCGGCGCCCAGATGCCCGGCCATGCCCCTCAGGGCGCCCCCGGGGA 261
        |||||| ||||||| ||  |||| |  ||| |||||  |  ||||| | |||||||| ||||
Db  187 CAGCCCCAGCCTGCCGCCTCTACTGCCCA------GCAATGCCCCTCCAGCCGCCCCGGGGA 239

Qy  262 CACCAGAGCAGCCTCGCTGCCGCGAGCCGGGAGGGAGGACCCAAAAGACGGAGGAGAGCAGGAGGA 321
        || ||||| ||| | |  | |  || | |||||| |||||||||||||||||||||||||||
Db  240 CACCAGCAGCTCCTCTCTCTGGAGCGGGAGGATGATCCGAAAGGAGGAGGAACAGGAGGA 299

Qy  322 GCCGCGGTTGGCAAGGGACGGAGGAGCGCGCAAAGACCCAGCACCACGACCGGGAGGTGGGGCGGC 381
        ||||||||| | ||||||  ||||||||| |||||||||||| ||||||| ||||||||| ||||
Db  300 GAACCGTTGGGCAAAGGAAGAGAGCGCGCCAGGAGCCCCGGTTGGAAAAGGCGGAGGTGGGGAGGCC 359

Qy  382 TGGGGAGGAAGCGCGCAAAGCGACCCCCCCGGTGGAAAAGCGGTGACACGCCAAAAGGACCCTGCGGT 441
        |||   |||||||  |||||||||||  |   ||| || || |  |  ||||  |  ||| |||| |
Db  360 TCGGCCGGAAAGCGGCAAGCACACCCACCGGTGGAAAGCAGTGACACCCCCAAAGGACCCAGCAGT 419

Qy  442 GATCTCCAAGTCCCATCCATGGCCCAGGAGACTCAGGCGCCCTCAGAGACTATTACCCAATGG 501
        ||  | ||||| | |  |  | |||||| ||||||||||| ||||||||||||||||||||
Db  420 GACCACCAAAGTCTCAGCCATGGCCCAGGACTCTGGCCCCTCAGATTCTGCTACCCAATGG 479

Qy  502 GGACTTGGAGAAGCGGAGTGAGCCCAGGAAGGGCCGAGAGAGGGAGAGGCCCTGCTGGGGGGCAGAA 561
        |||| |   || |||| |  ||||| ||||| | | | |  ||||| | |||||| ||||| ||||
Db  480 AGACTTGGAGAAGCGGAGTGAAACCCAACCTGAGGAAGGGGGTGAGCCCCAGCTGCAGGGCAA 539

Qy  562 GGGCGGGGCCCCCAGCAGAGGGAGAGGGTGCAGCTGAGACCCTGCCTGAAAGCCTCAAGAGC 621
        ||| || || || | | ||| | |   || |  |  | | ||| | | | ||| ||  | |
Db  540 GGGTGGGGCCCCAGCCTGAAAGGAGAGGG---AACTGAGACCCCACCAGAAGCCCTCCAGAGC 596

Qy  622 AGTGGAAAAATGGCTGCTGCACCCCCAAAGGAGGGGGCCGAGGAGCCCCTGCAGAAGCGGGGCAA 681
        |||| | | ||||||| |  | ||||||||||||| || || ||| | ||| ||||||||||||
Db  597 TGTGGAGAAATGGCTGCTGTTGTGTGACCAAAGGAAGCCGTGAGACCCTCGCAGGAGAGGGCAA 656

Qy  682 AGAACAGAAGGAGAGACCAACATCGAATCCATGAAAATGAAAATGAGAGGGTCCCGGGGGCTGCG 741
        |||||| | |   |  |   ||| |   |  |  ||  |    || |||||||||||||||||||
Db  657 AGAACAGAAGCAGACCAACATCGAATCCATGAAAAATGGAGGGTCCCGGGGGCCGCACTCGG 716

Qy  742 GGGTGGCTTGGGCTGGGAGTCCAGCCGTCCGTCAGGGGCCCATGCGAGGCTCACCTTCCA 801
        |||||  |||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  717 AGGTGGCTTGGGCTGGGAGTCCAGCCCTCCGTCAGGGGCACCCCATGCCAAGACTCACCTTCCA 776

Qy  802 GGCGGGGGGACCCCTACTACATCAGCAAGCGGAAGCGGGACGAGTGGCTGGCACGCTGGAA 861
        ||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db  777 GGCGGGGGGACCCCTACTACATCAGCAAGCGGAAGCGGGACGAGTGGCTGGCACGCTGGAA 861
```

Antibody / query matches (upper-left alignment columns, Qy/Db pairs):

```
Qy 2821 GGAAAGGGTATTTGGTTTCCCAGTCCAACTATACTGACGTGTCTCAACATGAGCCGCCTTGGC 2880
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 GGAAAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTGTCTCAACATGAGCCGCCTTGGC 2880

Qy 2881 GAGGCAGAGACTGCTGGGCCCGGCGGTGCCAGTCATCCGCCACCTCTTCGCTCC 2940
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 GAGGCAGAGACTGCTGGGCCCGGTCATGAGACCGTGCCAGTCATCCGCCACCTCTTCGCTCC 2940

Qy 2941 GCTGAAGGAGTATTTTGCGTGTGTGTGTGTGTTAAGGACAATGGGGGGCAAACTGAGGTAGCGACACA 3000
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 GCTGAAGGAGTATTTTGCGTGTGTGTGTGTTAAGGACAATGGGGGGCAAACTGAGGTAGCGACACA 3000

Qy 3001 AAGTT 3005
        |||||
Db 3001 AAGTT 3005
```

```
Qy   3018 ------AAAAAACACAAAACATAATAAAAACACCAAGAACATCAGGAGGATGGAGAAGTATCA 3072
                 ||||||||||||||||||||||||||||||||||||||||| ||||||| |
Db   2997 AAACAAAAACAACAAAACAACAATAAAAACACCAAGAACGAGAAAAGT-TCA 3055

Qy   3073 GCAACCAGAAGAGAGAAAAAGGAAATTTAAAAACAACCACACAGAGGGGGAAAATACCGGAGG 3132
                 ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   3056 GCAACCAGAAGAGAGAAAAAGGAAATTTAAAAGC--AAACCACAGAGGAGGGGGAGG 3113

Qy   3133 CTTTGCCTTGCGAAAAAGGGGTTGGACATCATCTCTGATTTTTCAATGTTATTCTTCAGTC 3192
                 |||- | |          ||||||||||||||||||||||||||| || |||
Db   3114 GCTTGGCCTTGCAAAAGGGTTGGACATCATCTCTGAGTTTTTCAAIIGTTAACCTTCAGTC 3173

Qy   4265 TACAAAACGT 4274
                 |||||||| ||
Db   4183 TACAAAATGT 4192
```

```
Qy   4087 TAACCCTTTGGATTGTTTTTCTAAAAGGAGAACTCCCTCGGCAAGAGATGGCAGAGAGGGGTACGGAG 4146
                 ||| ||||||||||||||||||||||| |||| ||||  ||||||        ||
Db   4009 AATAACCTTAGAATATGAATCTAAAATCGACTGCCCTCAGAAAAATGGC------TTGAGAA 4062

Qy   4147 TCTTCAGCGCCCAGTTTCTCCACTTTAGCCAAATTCGAGGGGCTCCTTGTGTGGGATCAGAAC 4206
                 | |||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4063 AACATTGTCCCTGATTTTGAATTCGTCAGCCACGTTGAAGGCCCCTTGTGGGGATCAGAAA 4122

Qy   4207 TAATCCAGAGTGTGGGAAAGTGACAGTCAA--AACCCACCTGGACAAATAAAAAAAACA 4264
                 |||||||||| ||||||||||||||||||  |||||||||||||||||||||||||||
Db   4123 TATTCCAGAGTGAGGGAAAAGTGACCCCGCCCATTAACCCGCCTGGACCAAAATAAAAAAAACA 4182
```

ACCESSION   BC007466
VERSION   BC007466.1 GI:13938620
KEYWORDS   MGC.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4094)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 8 Row: n Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6681208.

FEATURES   Location/Qualifiers
     source   1..4094
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="CZECH II"
                     /clone="MGC:5662 IMAGE:3492853"
                     /tissue_type="Mammary tumor metastatized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMTV."
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   245..2971
                     /codon_start=1
                     /product="DNA methyltransferase 3A"
                     /protein_id="AAH07466.1"
                     /db_xref="GI:13938621"
                     /db_xref="LocusID:13435"
                     /translation="MPSSGPGDTSSSSLEREDDRKEGEEQEENRGKEERQEPSATARK"

LOCUS   BC007466   4094 bp   mRNA   linear   ROD 07-AUG-2002
DEFINITION Mus musculus, DNA methyltransferase 3A, clone MGC:5662
            IMAGE:3492853, mRNA, complete cds.

RESULT 4
BC007466

```
Query Match   65.8%;  Score 2825.4;  DB 10;  Length 4094;
Best Local Similarity  84.1%;  Pred. No. 0;  Mismatches 546;  Indels 114;  Gaps 22;
Matches 3487;  Conservative 0;  Mismatches 546;  Indels 114;  Gaps 22;
```

```
VGRPGRKRKHPPVESSSDIPKDPAVTTKSQPMAQDSGPSDLLPNGDLEKRSEPQPEEGS
PAAGQKGPAPAEGEGTETPPEASRAVENGCCVTKERGASAGEGKEQKPTNIESMKME
GSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYIISRKRDEWLARWKREAEKKARVIA
VMNAVEENQASGESQKVEEASPAVVOQPTDPASPTVATTPEVGGDAGDKNATKAADD
EPEYEDGRGFGIGELVWKLRGFSWWMTGRSRAAEGTRNVWMFGDGKFSV
VCVEKLMPLSSPCSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFPACHDSDESDSG
KAVEVQNKQMIEWALGGFGPSGPRGLEPPEEEKNPYKEVYTDMVVEPEAAAYAPPPPA
KKPRKSTTEKPKVKEIIDERTRERLVYEYRQKCRNIEDICISCGGSLNVTLEHELFIGG
MCQNCNCFLECAYQDDDGYSYCTICCGGREVLMCGNNNCCRCRVECVDILVGPG
AAQAAIKEDPWNCYMCGHHGTYGLLRRREDWPSRLQMFFANNHDQEDPPKVYPPVPA
EKRKPIRVLSLEDGIAYGLLVLLDGIQVDRYIASEVCEDSITVGMYRHQGKIMYVGD
VRSVTQKHIQEWGPFDLVIGGSPCNDLSIYNPARKGLYEGTGRLFFEYRLLHDARPK
EGDDRPFWLFENVAMGVSDKRDISRFLESSNPVMIDAKEVSAAHRARYFWGNLPGMN
RPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQSKDQHFPVFMNEKEDILW
CTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV"
```

BASE COUNT   1067 a   1057 c   1164 g   806 t

ORIGIN

```
Qy      1 GCCGCGGCGCACCGAGGGGCGCGGCAGCCGGCCGGCCGACCCCCACCGGCCATACGGTGGCAGCC 60
          |||||||||| |||||| ||  | || |||||||||||||||||||||||| ||||  |  ||||
Db     26 GCCGCGGCGCAGGCGGAG----CCGGCGCGGCCGACCCCGCCGGCCCCACACGGCAGAGCC 81

Qy     61 ATCGAAGACCCCCACCCACAGGCTGACAGAGAGGCCACCGTTCACCAGAGGGCTCAACACCGG 120
          |||  |  |||  | |   ||||| |||||||||||||||||||||||||||||||||||
Db     82 GCCTGAACCCCAGCCCTGAGGCTG---------------CACTTTTCCGAGGCGCTTGACATCAGG 131

Qy    121 APCTATGTTTTAAGTTTTAACTCTCGCCTTCCAAAGACACCAGATAATTCCTTCCCAAAGCC 180
          ||||||||| ||||||||||||||||| |||||||||||||||||| ||||||||||||||
Db    132 GTCTATGTTTAAGTTCTTAGCTCTTGCTTACAAAGACCACGCAATTCCTTCTCTGAAGCC 191

Qy    181 C-AGCAGCCCCCCAGC-CCCGGCGCAGCACCCCAGCCTGCCTCCCGGCGGATGCCCGCC 238
          ||||||  ||||||  ||||| ||||||||||||||||||| ||||||||
Db    192 CTCGCAGCCCCACAGCGCCCTCGCCAGCCTGCCGCCCCTACTGCCCA------GCA 244

Qy    239 ATGCCCCTCCAGCGCGCGCCCCGGGGGAACCACCAGCACTCTGCTGCGGAGGAGGACCGA 298
          |||||| ||||||| |||||||||||||||||||||||||| |||||||||||||||
Db    245 ATGCCCTTCCAGCGCGCCCCGGGGGACACCAGCAGCTCCTTCTCTGAGCGCGGGGGAGATGATCGA 304

Qy    299 AAGGACGGAGAGGAGAGCAGCAGGAGGAGGAACCGTGGCCAAGGAGAGCCCCAGAGCCCACCACC 358
          ||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||||
Db    305 AAGGAAGGAGAGGAACCAGGAGGACCCAGAGCCCAGCGCC 364

Qy    359 ACGGCGACGGAAGGTGGGGGCGGCGGCCCTGGGGAGGAGAAAGCGGGT 418
          |||| ||||||||||||||||||| ||| ||||||||||||||||||||
Db    365 ACGGCCCGGAAGGTTGGGGAGGCCTGGCCGGAAGCGCAAGCACCACCGGTGGGAAAGCAGT 424

Qy    419 GACACGGCCAAAAGGACCCTGGGGTGATCTCCAAGTCCCCATCCATGGCCCGGACTCAGGC 478
          |||| ||| ||||||| ||||||||||| | |||||||||||||||||||||| ||||
Db    425 GACACCCCCAAAGACCCACCAGTGCACCAAGTCTCAGCCCAGGCCCAGGACTCTGGC 484

Qy    479 GCCTCAGAGCCTATTACCCCAATGGGGGACTTGGAGAAGCCGGAGTGAGCCCAGCCCAGGAG 538
          |||||||||||| || ||||  |||| ||||||||||||| |||||||||| ||||||||
Db    485 CCCTCAGATCTGCTACCCAATGGAGACTTGGAGAAGCGGAGTGAACCCCAACCTGAGGAG 544

Qy    539 GGGAGCCCTGCTGGGGGGGGCGGGGGGGCCCCAGCCACAGAGAGGGGGGCAGGGGGCTTC 598
          |||||||| |||| ||||||||||||||||||||| | || |||||||||||| |||
Db    545 GGGAGCCCAGCTGCAGGGCGCGGGGGGCCCAGCTGAAGGAGAGG----AACTGAC 601

Qy    599 ACCCTGCCTGAAGGCCTCAACAGCAGTGGGAAAAATGGGCTGCTGCTGCACCCCCAAGGAGAGGGCCC 658
          ||| |  ||||||||||||||| ||| | |||||| |  |  ||| ||||||||||||
Db    602 ACCCCACCAGAAGCCTCCAGAGCTGCGGGCTGTGTGCTCTGTGTGACCAAGGAGCCGT 661

Qy    659 GGAGCCCCTGCAGAAGCGGGGCAAAGAACAGAAGGAGACCAACATGAATCGAATCCATGAAAATG 718
          |||| |||||||||||||||||| ||| ||| ||||||||||||||| ||||||||||||||
Db    662 GGAGCCTCTGCAGAAGGCGGGCAAAAGAACAGAAGGAGACCAACATCGAATCCATGAAAAATG 721

Qy    719 GAGGGCTCCCGGGCCGCGGCTTCGGGGGTTGGGCTTGGGAGTTCCAGCGCTCAGCGG 778
          |||  | |||| ||||| ||||| ||||| || |  |  |  ||||||||| ||||
Db    722 GAGGGCTCCCGGGCCGCGGCTGCGAGAGGTGGCTTGGGAGTCCAGCGCTCCGTCAGCGA 781

Qy    779 CCCATGCCGAGGGTCCACCTTCCAGCGCCGGGGGACCCCTTACTACATCAGACAAGACGCAAGCGG 838
          |||||||| | ||  | |||||||||||||||||||||||||||||| ||||||||||
Db    782 CCCATGCCAAGACTCCACCTTCCAGCGGCGGGGACACCCTTACTACATCAGCAAACGGAAACGG 841

Qy    839 GACGAGTGGCTGGCCACGCTGGAAAAGGGAGGCTGAGAAGGAAAGCCAAGGTCATTGCAGGA 898
          |  |||||||||| ||||||||||||||||||||||||||||||||||||| ||||| |||
Db    842 GATGAGTGGCTGGCCACGTTGGAAAAAGGGAGGCTGAGAAGGAAAGCCAAGGTAATTGCAGTA 901

Qy    899 ATGAAATGCTGTGGAAGAAAACCAGGGGCGCCGGGGAGTGCTCAGAAGTGGAGGAGGCCAGC 958
          ||||||||||||||  ||  |||||||||| |||||||| |||||||||||||||||||||
Db    902 ATGAAATGCTGTGGAAGAGAAACCAGGCCTCTGGAGAGTCTCAGAAGGTGGAGGAGGCCAGC 961

Qy    959 CCTCCTGCTTGCTGTGCAGCAGCCCACGCAACCCGGGGAGTGCTTCTCCGACTGTGGCCACCACCCTGAG 1018
          |||||  |||| ||||| ||| || |||||||| |  |  |  |||||||||||||||||||||||
Db    962 CCTCCTGCTTGCTGTGCAGCAGCCCACGGACCCTGCTTCTCCGACTGTGGCCACCACCCTGAG 1021

Qy   1019 CCGGTGGGGGTCCCATGCTGGGGACAAGAAAGTGCCACAAGCAGGGAGAGATGAGACCGAGAG 1078
          | ||| || ||  | |||||||||||||||| ||| |||||||||||||||||||||||||
Db   1022 CCAGTAGGAGGGAGGGATGCTCGGGGACAAGAATGCTACCAAAGCAGCCGGAGCCTGAG 1081

Qy   1079 TACGGAGGACGGGCCCGGGTCGGGGCTTTGGCATTGGGGGAGGCTGGGTGTGTGGGGGGAAACTTCGGGGCTTC 1138
          | |||| | ||||||||||| || | ||||| ||||||||||| || || ||||||||||||
Db   1082 TATGGAGGATGGGCCCGGGGTCGGGGCTTTGGCATTGGGGGAGGCTGGGTGTGTGTGGGGGGGCTTC 1141

Qy   1139 TCCTGGTGGCCAGGCGCGCCCATTGTGTCGTTGGTGGATGGACGGGGCCGGGGCCGAGCCACTGAA 1198
          |||||||| |||  ||||| ||||| | |  ||||| ||  || |||||||||||||||||||
Db   1142 TCCTGGTGGCCAGGCGCGAATTGTGTCGTTGGTGGATGACAGGGCCGGGCAGCTGAA 1201

Qy   1199 GGCACCCGCTGGGGTCGTTCATGTGGTTCGGAGACGGGCAAATTCTCAGTGGTGTGTTGAGAAG 1258
          |||||| ||  |  |||||||| | | ||||||||||||||||||||||||||||||||||
Db   1202 GGCAGCTCGGCGGTCAGTGCTGGGTCAAGTGTGGTCCTGGGCAGCCAGGCAGGCGCTG 1261

Qy   1259 CTGATGCGCGCTGAGCTCGTTTTGCAGTGCGTTCCACCAGGGCCACGTGTGTCAGGGAAGCTG 1318
          || ||||||||  || |||| |||||||| |||||||||||||| ||||||||||||||||
Db   1262 CTCATGCCGCTGAGCTCCTTCTGCAGTGCAATTCCACCAGGGCCACCTACAAACAGCCC 1321

Qy   1319 ATGTACCGGCCAAAGCCATCTACGAGGTCCTGCAGGTGGCCAGCAGCCGCGCGGGGAAGCTG 1378
          |||||| |||||||| |||||||||| ||||||||||||| |||||||||||||||| |||
Db   1322 ANGTACCGCCAAAGCCATCTACGAAGTCCTCCCAGGTGGCAGCAGTTCTCAGGTGTGGGAAAGCTG 1381

Qy   1379 TTCCCGGTTGTGCCACGACCACCGGATAGAGTGGACACTGCCAAGGGCCGTGGAGGTTGCAGAAG 1438
          ||  ||||| ||||| ||  | |||| |||| |||||||||||||||||||||| ||||||
Db   1382 TTTCCAGCTTGCCCATGACAGTGAGTGAGGCAGAAAGTGACAGTGGCAAGGCTGTGTGGAAGTGCAGAAC 1441

Qy   1439 AAGCCCATGAATTGAATGGGGGCCCCTGGGGGCGGGCCTTCTGGGCCCCTAAGGGCGGCCTGGAG 1498
          ||||  ||| || ||||||||||||||||||| |||||||||||||||| |||
Db   1442 AAGCAGCATGATTTGAATGGGGCCCTGCGCTTCAACCGGCTGGCTTTCCAGCCGGGCGGGAA 1501

Qy   1499 CCACCGAAGAGAAGAGAGAAATCCCTACAAAGAAAGTGTACACGGACACATGTGGGGTGGAACCT 1558
          |||||||| |||||||||| |||||| ||||||||| ||||||||||||||||||||||||
Db   1502 CCACCAGAAGAGAAGAAGAATCCTTACAAAGGAAGTTTACACCGACATGTGGGTGGACCT 1561

Qy   1559 GAGGCAGCTGCCTACGGCACCACCTCCACGGCCCAGCCCAGGGAG 1618
          |||| ||||||||| | ||| ||  |  |||| || || |||||
Db   1562 GAAGCAGCTGCCTACGGCACCACCTCCACGGCCCAGCTCGGGCAG 1621

Qy   1619 AAGCCCCAAGGTCAAGGAGATTATTGATGAGCGGCCACAAGAGAGGGCGGCTGGGGTGTGTACGAGGTG 1678
          || || || ||||||||| |||||| || ||||| ||||||||||||||||| ||||| |||
Db   1622 AAACCTAAGGTCAAGGAGATCATTGACGAGCGGACCGAACGGGAGCGGCTGGTGTATGAGTATAGAGGTG 1681

Qy   1679 CGGGAGAGTGCCGGAACATTGAGGACATCTGCATCTCCTGTGGGGAGTCGCTTAACGTTACC 1738
          ||| || ||||||| ||||||| || ||||| || |||||||| |||||| |  |||||||
Db   1682 CGCCAGAAGTGCAGAAACATCGAGGACATTTGTATCTCATGTGGGAGGCTCAAATGTACC 1741

Qy   1739 CTGGAAACACCCCCTCTTCGTTGGACGGAAATGTCCCAAAAACTGCAAGAAACTGCTTTTCTGGGAG 1798
          |||| |||| ||  ||||  ||||| ||  ||||||||||| ||||||||||||| ||||| ||
Db   1742 CTGGACACCACCCACTCTTCATTGGAGGCCATGTGCCAGAACTGTAAGAACTGCTTTCTGGAG 1801

Qy   1799 TGTGCGTACCAGTACGACGACGGCTACCAGTCCTACTGCACCATCTGCTGTGGGGGC 1858
          |||||||| | || ||||||||  ||||| || |||||||| ||||| || |||||
Db   1802 TGTGCTTACCAGTATGACGACGGATACTCCTATTGCACCATCTGCTTGTGGGGGG 1861
```

```
Qy  1859 CGTGAGGTGGCTCATGTGCGGGAAACAACAACTGCTTGCTGCAGGTGGTTGTTGCGTGAGTGTGTG 1918
Db  1862 CGTGAAGTGCTCATGTGGTGGGAACAACAACTGCTCGCAGGTGGTTTGTGTCGAGTGTGTG 1921
         |||||||||||||||||||||||||||||||||||||||||| | ||||||||||

Qy  1919 GACCCTCTTGCGTGGGGGGCCCGGGGGGCTGCCCAGGCAGCCATTAAGGAAGGAAGACCCCTGGAACTGC 1978
Db  1922 GATCTCTTGGGTGGGGGGCACCAGGAGCTGCTCAGGCAGCAGCCATTAAGGAAGACCCCTGGAACTGC 1981
         || ||||||| ||||||| |||||

Qy  1979 TACATGTGCGGGCACAAGGGTACCTTACGGGGGTACTTGGGCCCTCC 2038
Db  1982 TACATGTGCGGGCATAAGGGCACCTTATGGGGTCGCCTGCCAAAGACGGGAAAGACTGGCCTTTCT 2041

Qy  2039 CGGCTCCAGATGTTCTTCCGCTAATAACCACGACCAGGAATTTGACCCTCCAAAGGTTTAC 2098
Db  2042 CGACTCCAGATGTTCTTTTTGGCCAATAACCATGACCAGGAATTTCGACCCCCAAAAGGTTTAC 2101

Qy  2099 CCACCTGTCCCAGCTGAGAAGAAGAAGAGAAGCCCATCCGGGTGCTGTCTCCTCTTTGATGGAATC 2158
Db  2102 CCACCTGTGCCCAGCTGAGAAGAGGAAGCCCATCCGGGTGCTGTCTCTCTTTGATGGGATT 2161

Qy  2159 GCTACAGGGGCTCCCTGGTGCTGAAGGAGGACTTCAGGTGGACCCGTACCATTGCCCTCG 2218
Db  2162 GCTACAGGGGCTCCCTGGTGCTGAAGGACCTGGGGCATCCAAAGTGGACCCGTACCATTGCCTCC 2221

Qy  2219 GAGGTGTGTGAGGGACTCCCATCACACGGTGGGCCATGGGGGCGCCACCAGGGGGAAGGAATCATGTAC 2278
Db  2222 GAGGTGTGTGAGGGACTCCCATCACACGGTGCGGCCACCAGGGGAAAAGATCATGTAC 2281

Qy  2279 GTCGGGGACGTCCGCAGCGTCACACACAGAGAAGCAATATCCCAGGAGTGGGGGCCCATTCGATCTG 2338
Db  2282 GTCGGGGAACGTCCCGGACAGCGTCACACAGAAAGCAATATCCAGGAGTGGGGGCCCATTCGACCTG 2341

Qy  2339 GTGATTGGGGGGCAGTCCCCTGCAAATGAACCTCTCCATCGTCAAACCTGCTCGCAAAGGGCCTC 2398
Db  2342 GTGATTGGAGGCAGTCCCTGCAAATGAACCTCTCAATTGTCAAACCTGCCCGCAAAGGGACTT 2401

Qy  2399 TACGGAGGGCACTGGGCCGCCGCTCTTCTTTTCTTTGAGTTCTACCGCGCCTCCCGCGGGCCCC 2458
Db  2402 TATGAGGGTACTGGCCGCGCCCTTCTTCTTTTTGACTTCTACCGCCATGCGGGCGGCCC 2461

Qy  2459 AAGGAGGGGACAATGGATCGCCCCTTCTTCTTCTGGCTCTTTTGAGAATGTGGTGGTGGCCATTCGACTGC 2518
Db  2462 AAGGAGGGGACGATGAATCGCCCCTTCTTCTTTTGAGAAATGTGGTGGCCATGGGGCGTT 2521

Qy  2519 AGTGACAAGAGCAGCACATCTCGGCGATTTTCCTGCTCTTCGAGGTCCACACCCCTGTGCATCGACCAAAA 2578
Db  2522 AGTGACAAGAGGGACAATCTCGCGATTTCTTCGAGTCTAAGCCCGTGATGATTGACGCCAAA 2581

Qy  2579 GAAGTGTCAGTGCACACAGGGCCCCCCTACTTCTTGGGGTAACCTTCCCGGTATGAAACAGG 2638
Db  2582 GAAGTGTCTGCTGCACACAGGGCCCGTTACTTCTCTGGGTTAACCTTCCTGGCCATGAAACAGG 2641

Qy  2639 CCGTTGGCATCCACTGTGAAATGATGAAGCTGCAGGAGCTGCACAGGAGCTGTCTCAGGCAGCATGGCAGG 2698
Db  2642 CCTTTGGCATCCCATGTGAATGATAAAGCTGCACGAGTGTCTGAGCACGGCAGA 2701

Qy  2699 ATAGCCAAAGTTCAGCAAAAGTGAGGAGGACCATTTACTACGAAGGTCAAACTCCATAAAAAGCAGGGC 2758
Db  2702 ATAGCCAAGTTCAGCAAAAGTGAGGACCATTACCACGAGGTCAAAACTCTATAAAAGCAGGGC 2761

Qy  2759 AAAGACCAGCATTTTCCTGTCTTCATGAATGAGAAAGAGGACATCTTTATGGTGGCACTGAA 2818
Db  2762 AAAGACCAGCAATTCCCCGTCTTCATGAACGAGGAGGACATCCTGTGTGGCACTGAA 2821

Qy  2819 ATGGAAAAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTCTCCAACATGAGCCGGCTTG 2878
Db  2822 ATGGAAAAGGGTGTTTGGCTTTCCCCGTCCACTACACAGACGTCTCCAACATGAGCCCGCTTG 2881

Qy  2879 GCGAGGCCAGAGAGAGACTCCTTGGGCCCGGTTCATGGGAGGCCGCCACCTTCTTCCCT 2938
Db  2882 GCGAGGCAGAGAGACTGCTTGGGCCGGCCGATCGTCATCGCCACTCTTTCCGCT 2941

Qy  2939 CCGCTGAAGGAGTATTTTGCGTGTGTGTAAGGGACAATGGGGGCAAAACTGAGGTAGGCGA-- 2996
```

```
Db  2942 CCGGCTGAAGGAGAATATTTTGCTTGTGTGTGAAGGGACAATGGGGGCAAAACTGAAGTAGGTGATG 3001
         |||||||||||||||||||||||||||||||||||||||||||||||

Db  2997 --CACAAAGTTAAACAAACAAAC------AAAAACAACAAAACATTAAATAAAAACACCAAGAA 3049

Qy  2997 ATAAAAAAGTTAAACAAACAAAACAAAACAAAACAACAAAACAACCAAGAA 3061

Db  3002 CATGAGGATGGAGAGAAGTAATCACGCACCCCAGAAGAGAGAAAAAGGAATTTTAAAACAAAAAACC 3109

Qy  3050 CGAGAGGACCGGAGAAAAGT-TCAGCACCCAGAGAGAGAAAAGGAATTTAAAGC--AAACC 3118

Db  3062 ACAGAGGCGGAAAATACCGGAGGGCGTTTGCCTTGCCGAAAAAGGGTTTGGGACATCAATCTCCTGA 3169

Qy  3110 CCCCTTCCCTTTTTTTGGTCAGACCTTTTATTTTCTCACTC-TTTTCAGGAGGGGTTTTTC 3288

Db  3119 ACAGAGGAGGAAAAACGCCCGGAGGGCTTGGCCTTGCCAAAAAGGTTGGACATCAATCTCCTGA 3178

Qy  3230 CCCCTTCCCTTTTTTTTCGGTCAGACCTTTTATTTTCTCACTC-TTTTCAGGAGGGGTTTTTC 3288

Db  3238 CCC---TCCGGTCCTAGGAGGCGAACTTTTTGTTTTTCTACTCCTTTTTTCAGGAGGGGTTTTTC 3294

Qy  3289 TGTTTGTTTTGGG--TTTTTGTTTTTCTCTGTGTGACTGAAACAAAGAAGGTTTAATTGCAGCAAAAA 3347

Db  3295 TGTTTGTTTGGGGTTTTTGTTTTTTTCTTGTCCTGGTGACTGAAACAAGAGAGTTATTGCAGC-AAAA 3353

Qy  3348 TCAGTAACAAAAAATAGTAACAAATACCTTGCAGAGGAAAAGTGGGAGGAGGAGAGAAAAAAG 3407

Db  3354 TCAGTAACACAAAAAGTAGTAGCCTTG----------GAGAGAAAAGGGAGGAGAGAG 3399

Qy  3408 GGAAAATTTTTAAAAGAAAAATCTATATATATTTGGGTTGTTGTTTTTTTGTTTTTGTTTTTTTTTT 3467

Db  3400 GGAAAAANTTCTATAAAAAACTTAAAATTAATTGGGTTTTTTTTTTTTCCCTTTTC----------- 3449

Qy  3468 TTTTGGGTTTTTTTTTTTTTTTTTTTACTATATATATCTTTTTTTTTTTTGTTTGTTCTCTAGCCTGATCAGATA 3527

Db  3450 --------------------------------TATATATATCTCTTTGGTTTGTTCTCTAGCCTGATCAGATA 3486

Qy  3528 GGAGCACAAGCAAGGGGACGGGAAAAGAGAGAGAGACACTCAGGCGGCAGCAGCAATTCCCTCCCAGCC 3587

Db  3487 GGAGCACAAAACAGGAAGAGAAT---AGAGAACCCTCGGAGGCAGAGTGCTCCTCTTCCCACC 3542

Qy  3588 ACTGAGCGTGTCCGTGCCAGCACCCATTCCCTGGTCACGCAAAACAGAACCCAGTTAGCAGCAG 3647

Db  3543 CCCGAGCGAGTCTCAAACAGCCACCATTCCTGGTCATCGCAAAACAGAACCCAACTAGCAGCAG 3602

Qy  3648 GGAG-- ACGAGAAACACCACACAAAGACACATTTTTTTTTCTACAGTATTTTTCAGGTGGCCTACCACACA 3705

Db  3603 GGCGCTGAGAGAACACACCACACCAGACACTTTTCTACAGTATTTTCAGGTGGCCTACCACACA 3662

Qy  3706 GGAAAACCTTGAAGAAAAATCACGTTTTCTAGAAGGCCCGCTGTGTTACCTCTTGTTTACAGTTTATA 3765

Db  3663 GGAAAACCTTGAAGAAAAACCAGTTTCTAGAAAGCCCGCTGTTTACCTCTTGTTTACAGTTTATA 3722

Qy  3766 TATATATAGATAGGATATGAG--ATATATATATAAAAGGTACTGTTAAACTACTGTACATCC 3823

Db  3723 TATATATGATAGATAGTATGAG---ATATATATATAAAAGGTACTGTTAAACTACTGTACATCC 3782

Qy  3824 GACTTCATAAATGGTGCTTTC-AAAACAGCGAGAGTGAGTAAAAACAACATCAGCGTTCCACGTTGC 3882

Db  3783 GACTTCAATAAATGGTGCTTTTCAAAAACAGCGAGATGAGCAAAGACAATCAGCTTCCCCTGGC 3842

Qy  3883 CTTCTCGCGCAAAAGGGTTTTCACCAAGGATGGAGAAAAGGGACACACTTGCAGATGGCGGT 3942

Db  3843 CCTCTGTTGCAAAAGGGTTTCAGCCCAGGATGGGGGAGGGGAGCAGCTGGAGGGGGTTTTA 3902

Db  3903 ACAAACTGAAGGATGACCCATATCACCCCCCAC---------CCCTGCCCCATGCC 3949

Qy  3943 TCTCACGGTGGGCGCTCTTCCCCTGGTTTTGTTAACGAAGTGAAGTGGTCGGCCCGGTGGGACAGGCTGAGAG 4002

Qy  4003 AGGTTCTCCCTGCCAAAAAGGGGGCTAGATGAGGTGGTCGGCCCGTGGGACAGCTGAGAG 4062
```

```
RESULT 5
AK025230

LOCUS       AK025230                2191 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ21577 fis, clone COL06724, highly similar to
            Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA.
ACCESSION   AK025230
VERSION     AK025230.1  GI:10437699
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL06724.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikena,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2191)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..2191
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COL06724"
                     /tissue_type="colon"
                     /clone_lib="COL"
                     /note="cloning vector pME18SFL3"
     misc_feature    1..2191
                     /note="highly similar to AF067972 Homo sapiens DNA
                     cytosine methyltransferase 3 alpha (DNMT3A) mRNA"
BASE COUNT    571 a    505 c    575 g    540 t
ORIGIN
```

```
Query Match          49.8%;  Score 2139.6;  DB 9;  Length 2191;
Best Local Similarity  99.4%;
Matches 2169;  Conservative 0;  Mismatches 0;  Indels 4;  Gaps 2;
```

```
Qy   3950 TAGCTTCACCTGCCAAAAAGGGGCTCAGCTTGGGACCCTGGGGAAGCTGAGTG 4009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4063 TGGGATTCATCCAGACTCATGCAATAACCCTTTGATTGTTTTCTTAAAAGGAGACTCCCTC 4122

Qy   4010 TGGAATTTATCCAGACTCGCGTGCAATAACCTTAGAAATATGAATCTAAAATGAACTGCCTC 4069
          |||||| |||||||||| | |||  |  |   ||  || |||  || ||||||||||||||
Db   4123 GGCAAGA 4129

Db   4070 AGAAAAA 4076
```

```
Qy    121 TGGAGGAAATGTGCCAAAAACTGCAAGAACTGCTTTCTGGAGTTGCGGTACCAGTACGAGGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TATTGATGAGCGCACAAGAGAGAGCGGGCTGGTGTACCAGGTGCCGGAACAT 60

Qy    181 CGACGGCTACCAGTCCTTACTGCACCATCTGCTGTTGTGGGGGCCGTGAGTGTGCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGAGGACATCTGCATCTCCTGTGTGGGAGCCTCAAGTGTTACCCTGGGACAACCCCTCTTGT 120

Qy    241 AAACAACAACTGCTGCAGGTGCTTTTGCCGTGGAGTGTGTGGGACCCTCTTTGGTGTGGGCCGGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTCATGAATGAGAAAGAGGACATCTTATGGTGCACTGAAATGTGGAAAGGGTATTGGGTTT 180

Qy    301 GGCTGCCCAGGCAGCCATTAAGGAAGAACCCCTGGAACTGCTACATGTGCGGCACAAGGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGACGGCTACCAGTCCTTACTGCACCATCTGCTGTTGTGGGGGCCGTGAGTGTGCGG 240

Qy    361 TACCCTACGGGCGTGCTTGCCGTGGGCGGCGGAGAGGAGACTGGCGCCCCGGCTCCCAGACGTTCTTCGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAACAACAACTGCTGCAGGTGCTTTTGCCGTGGAGTGTGTGGGGACCCTCTTTGGTGTGGGCCGGG 300

Qy    421 TAATAAACCACGACCAGGAAATTTGACCCTCCAAAGGTTTTACCCACCTGTTCCCAGCTGAGGA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCTGCCCAGGCAGCCATTAAGGAAGAACCCCTGGAACTGCTACATGTGCGGCACAAGGG 360

Qy    481 GAGGAAGCGCCATCCGGGTGCTGGTGTCTCTCTTTTGATGGAATGCCTACACGTGCTCCTGGTTCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TACCCTACGGGCGTGCTGCTGCCGTGGGCGGCGGAGAGGAGACTGGCGCCCCGGCTCCCAGACGTTCTTCGC 420

Qy    541 GAAGGACTTGGGGCATTCAGGTGGACCGGCTACATTGCCCTCGGGAGGTGTGTGAGGACTCCAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TAATAAACCACGACCAGGAAATTTGACCCTCCAAAGGTTTTACCCACCTGTTCCCAGCTGAGGA 480

Qy    601 CACGGTGGGCATGGTGGGCCACCAGGGGAAGAATCATGTAGTCTGGGACGTGCCGCACGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGGAAGCGCCATCCGGGTGCTGGTGTCTCTCTTTTGATGGAATGCCTACACGTGCTCCTGGTTCG 540

Qy    661 CACACAGAAAGCAATATCCAGGAGTGGGGGCCCATTCGATCTCGGTGATGGGGGCAGTCCCTG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAAGGACTTGGGGCATTCAGGTGGACCGGCTACATTGCCCTCGGGAGGTGTGTGAGGACTCCAT 600

Qy    721 CAATGACCTCTCCATCGTCAACCCTGCTCGGCCTGCAAGGGCACTGCCGGCGT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CACGGTGGGCATGGTGGGCCACCAGGGGAAGAATCATGTAGTCTGGGACGTGCCGCACGT 660

Qy    781 CTTCTTTGAGTTGTTCTACCGCCTCCTGCACATGAATGCGCCCCAAGGAGGAGGAGATGATCGCCC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACACAGAAAGCAATATCCAGGAGTGGGGGCCCATTCGATCTCGGTGATGGGGGGCAGTCCCTG 720

Qy    841 CTTCTTTGCTGGCTCTTTGAGAAATGGTGGCCATGGGCGTGGTTAGTGACAAGGAGGGACATCTC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAATGACCTCTCCATCGTCAACCCTGCTCGGCCTGCAAGGGCACTGCCGGCGCT 780

Qy    901 GCGATTTCTCGAGTCCAACCCTGTGATGATTGATGCCAAAGAAGTGTCAGCTGCACACAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTTCTTTGAGTTGTTCTACCGCCTCCTGCACATGAATGCGCCCCAAGGAGGAGGAGATGATCGCCC 840

Qy    961 GGCCCCCTACTTCTCTGGGGTAACCTTCCCGGTATGAACAGGCCGTTGGCATCACTGTGAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTTCTTTGCTGGCTCTTTGAGAAATGGTGGCCATGGGCGTGGTTAGTGACAAGGAGGGACATCTC 900

Qy   1021 TGATAAGCTGGAGCTGCAGGAGTGTCTCAGGAGCATGGCAGGATAGCCAAGTTCAGCAAAGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GCGATTTCTCGAGTCCAACCCTGTGATGATTGATGCCAAAGAAGTGTCAGCTGCACACAG 960

Qy   1081 GAGGACCATTACTACGAGGTCAAAACTCCATAAAGCAGGCGAAAGAGACCAGCATTTTTCCTGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GGCCCCCTACTTCTCTGGGGTAACCTTCCCGGTATGAACAGGCCGTTGGCATCACTGTGAA 1020

Qy   1141 CTTCATGAATGAGAAAGAGGACATCTTATGGTGCACTGAAATGTGGAAAGGGTATTGGGTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TGATAAGCTGGAGCTGCAGGAGTGTCTCAGGAGCATGGCAGGATAGCCAAGTTCAGCAAAGT 1080

Qy   1201 ... 
```

```
Qy   1639 TATTGATGAGCGCACAAGAGAGAGCGGGCTGGTGTACCAGGTGCCGGAACAT 1698
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TATTGATGAGCGCACAAGAGAGAGCGGGCTGGTGTACCAGGTGCCGGAACAT 60

Qy   1699 TGAGGACATCTGCATCTCCTGTGTGGGAGCCTCAAGTGTTACCCTGGGACAACCCCTCTTCGT 1758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGAGGACATCTGCATCTCCTGTGTGGGAGCCTCAAGTGTTACCCTGGGACAACCCCTCTTGT 120

Qy   1759 TGGAGGAAATGTGCCAAAAACTGCAAGAACTGCTTTCTGGAGTTGCGGTACCAGTACGAGGA 1818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTCATGAATGAGAAAGAGGACATCTTATGGTGCACTGAAATGTGGAAAGGGTATTGGGTTT 180
```

```
RESULT 6
BC018214
LOCUS        BC018214              1758 bp    mRNA    linear   PRI 06-DEC-2001
DEFINITION   Homo sapiens, clone IMAGE:3862699, mRNA, partial cds.
ACCESSION    BC018214
VERSION      BC018214.1  GI:17390481
```

KEYWORDS   .
SOURCE   Homo sapiens
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1758)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-DEC-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
COMMENT   Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Baylor College of Medicine Human Genome
             Sequencing Center
             Center code: BCM-HGSC
             Web site: http://www.hgsc.bcm.tmc.edu/cdna/
             Contact: villalon@bcm.tmc.edu.
             Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
             A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
             Muzny,D.M., Gibbs,R.A.
             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 21 Row: h Column: 23.
FEATURES   Location/Qualifiers
     source   1..1758
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3862699"
                     /tissue_type="Ovary, adenocarcinoma"
                     /clone_lib="NIH_MGC_66"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   <1..859
                     /codon_start=2
                     /product="Unknown (protein for IMAGE:3862699)"
                     /protein_id="AAH18214.1"
                     /db_xref="GI:17390482"
                     /translation="AEKKRPIRVLSLFPGIAYCLLVLKDLGIQVDRVIASEVCEDSIT
                     VGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYEGTGR
                     LFEFYRLLHDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSA
                     AHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTTTRSNSIKQGRD
                     QHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFA
                     PLKEYFACV"
BASE COUNT   513 a   376 c   428 g   441 t
ORIGIN
```

Query Match   38.8%; Score 1667; DB 9; Length 1758;
Best Local Similarity 99.2%;
Matches 1697; Conservative 0; Mismatches 10; Indels 4; Gaps 2;

```
Qy   2110 AGCTGAGAAGAGGAAGCCCCATCCGGGCGTGCTCTCTCTCTTTGATGGAATCGCTACAGGGCT 2169
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCTGAGAAGAGGAAGCCCCATCCGGGCGTGCTCTCTCTCTTTGATGGAATCGCTACAGGGCT 60

Qy   2170 CCTGGTGCTGAAGGACTTGGGCATTCAGGTGGACCGCGTACATTGCCTCGGGAGGTGTGTGA 2229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTGGTGCTGAAGGACTTGGGCATTCAGGTGGACCGCGTACATTGCCTCGGGAGGTGTGTGA 120

Qy   2230 GGACTCCATCACGGTGGGCATGGTGCGGCACCAGGGAAGAGATCATGTACGTCGGGGACGT 2289
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGACTCCATCACGGTGGGCATGGTGCGGCACCAGGGAAGAGATCATGTACGTCGGGGACGT 180

Db    181 CCGCAGCGTCACACAGAAGCATATCCAGGAGTGGGGGCCCATTCGATCTGGTGATTGGGGG 240
```

Db   1201 CCCAGTCCCACTATACTGACGTCTCCAACATGACCGGCGCTTGGCGAGGCGAGAGGACTGCTGGG 1260
Qy   2899 CCGGTCCATGGAGCGGTGCCAGTCAATGCATCCGGCCCACCCTCTTCGCTCCCCTGAAGGAGTATTTTGC 2958
Db   1261 CCGGTCCATGGAGCGGTGCCAGTCATCCGGCCACCTCTTCGCTCCCCTGAAGGAGTATTTTGC 1320
Qy   2959 GTGTGTGTGTAAGGGACGCATGGGGGGCAAAACTGAGGTAGCGACACAAAGTTAAACAAACAAACA 3018
Db   1321 GTGTGTGTAAGGGACGCATGGGGGGCAAAACTGAGGTAGCGACACAAAGTTAAACAAACAAACA 1380
Qy   3019 AAAAACACAAAACATAATAAAAACACCAAGAACATGAGGATGGAGGAAGAAGTTAAACAAACAAAACA 3078
Db   1381 AAAAACACAAAACATAATAAAAACACCAAGAACATGAGGAAGAAGTTATCAGCACCC 1440
Qy   3079 AGAAGAGAAAAAGGAAATTTAAAACAAAAAACACACAGAGGCGGGAAAATACCCGGAGGGCTTTGC 3138
Db   1441 AGAACAGAAAAAAGGAAATTTAAAACAAAAAACACACAGAGGCGGGAAAATACCCGGAGGGCTTTGC 1500
Qy   3139 CTTGGCGAAAAGGGTTGGGACATCATCTCCTGAATTTTCAAATGTTAATTCTTCAGTCCTCTATTT 3198
Db   1501 CTTGGCGAAAAGGGTTGGACATCATCTCCTGAATTTTCAAATGTTAATTCTTCAGTCCTTATTT 1560
Qy   3199 AAAAAACAAAAACCAAGCTCCCTTCCCTTCCCTCCCCTTTTTTTTTTTCGGTCAGACCTT 3258
Db   1561 AAAAACAAAAACCAAGCTCCCTTCCCTTCCCTCCCCTTTTTTTTTTTTCGGTCAGACCTT 1620
Qy   3259 TTATTTTCTACTCTTTTCACAGAGGGGTTTTTCTCTGTTTGTTTTGTTTGTTTTGTTTGCTCGTGA 3318
Db   1621 TTATTTTCTACTCTTTTCACAGAGGGGTTTTTCTCTGTTTGTTTTGTTTGTTTTCTTGCTCGTGA 1680
Qy   3319 CTGAAACAAGAAAGGTTATTGCAGCAAAAAATCACTATAACAAAAAATAGTAACAATACCTTGC 3378
Db   1681 CTGAAACAAGAAAGGTTATTGCAGCAAAAAATCACTATAACAAAAAATAGTAACAATACCTTGC 1740
Qy   3379 AGAGGAAAAGGTGGGAGGAGGAGGAAAAAAGGGAAAATTTTTAAAGAAAAATCTATATATTGGGT 3438
Db   1741 AGAGGAAAAGGTGGGA-GAGAGGAAAAAAGGGAAAATTCTATA---GAAATCTATATATTGGGT 1796
Qy   3439 TGTTTTTTTTTTTTTTTTGTTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTACTATATATCTT 3498
Db   1797 TGTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTGGGTTTTTTTTTTTTTTTTTTACTATATATCTT 1856
Qy   3559 CACTCAGGCGGCGGGCAGCCAATTCCCTCCCCAGCCCACTGAGCTGTCGTGCCCAGCCACCATTCCTCGGT 3618
Db   1857 TTTTTTGTTGCTCTCTAGCCTGATCAGATAGGAGCACACAAGCAGGGGACGGAGAAAGAGAGA .1916
Qy   3619 CACGCAAAAACAGAAACCCAGTTAGCAGCGACGACGAGGAGGAGAACACCACACAACGACATTTTTTC 3678
Db   1917 CACTCAGGCGGCGGGCAGCAATTCCCTCCCCAGCCCACTGAGCTGTCGTGCCCAGCCACCATTCCTCGGT 1976
Qy   3679 TACAGTATTTCAGGTGCCTACCACACAGGAGAAACCTTGAAGAAAAATCAGTTTCTAGAAGCC 3738
Db   1977 CACGCAAAAACAGAAACCCAGTTAGCAGCGACGACGAGGAGGAGAACACCACACAAGACATTTTTTC 2036
Qy   3739 GCTGTTTTACCTCCTTGTTTGTTTTACAGTTTTATATATATAATATGAGATATATATATATAAAA 3798
Db   2037 TACAGTATTTCAGGTGCCTACCACACAGGAGAAACCTTGAAGAAAAATCAGTTTCTAGAAGCC 2096
Qy   3799 GGTACTGTTTAACTACTGTACAA 3820
Db   2097 GCTGTTTTACCTCCTTGTTTGTTTTACAGTTTTATATATATAATATGAGATATATATATAAAA 2156
Db   2157 GGTACTGTTTAACTACTGTAAAA 2178

```
Qy   2350 CAGTCCCTGCAATGACCTCTCCATCGTCAAACGTCTGCGAAAGGGCCTCTACGAGGGCAC 2409
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGTCCCTGCAATGACCTCTCCATCGTCAAACGTCTGCGAAAGGGCCTCTACGAGGGCAC  300

Qy   2410 TGGGCCGGCTCTCTTCTTTGAGTTCTTACCGGCTCCGCTCCTGCATGATGCCGGGAGA   2469
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGGGCCGGCTCTCTTCTTTGAGTTCTTACCGGCTCCGCTCCTGCATGATGCCGGGAGA    360

Qy   2470 TGATCGCCCCCCTCTTCTTCTGGCCTCTTTGAGAAGTGTGGCCGTTAGTGACAAGAG   2529
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGATCGCCCCCCTCTTCTTCTGGCCATGGCCATGGCCGTTAGTGACAAGAG         420

Qy   2530 GGACAGCATCTCGGCGATTTTCTCGAGTCCAACCCTGTGATGATTGAATGCCAAGAACTCACC 2589
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGACAGCATCTCGGCGATTTTCTCGAGTCCAACCCTGTGATGATTGAATGCCAAGAACTCACC  480

Qy   2590 TGCACACAGGGCCCGGCTACTTCTTCTTGGGGGTAACCTTCCCGGTATGAACAGGCCCGTTGGCATC 2649
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCACACAGGGCCCGGCTACTTCTCTGGGGTAACCTTCCCGGTATGAACAGGCCCGTTGGCATC  540

Qy   2650 CACTGTGAAATGATAAAGCTGGAGTGCAGGAGCATGGGCAGGATAGCCCAAGTT      2709
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CACTGTGAAATGATAAAGCTGGAGTGCAGGAGCATGGGCAGGATAGCCCAAGTT       600

Qy   2710 CAGCAAAAGTGAGGACCAATTACTACGAGGTCAAACTCCATAAAGCAGGGCAAAGACCAGCA 2769
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CAGCAAAAGTGAGGACCAATTACTACGAGGTCAAACTCCATAAAGCAGGGCAAAGACCAGCA  660

Qy   2770 TTTTTCCTGGTCTTCATGAATGAGAAAAGAGGACATCTTATGGGTGCCACTGAAATGAAAGGGT 2829
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTTTTCCTGGTCTTCATGAATGAGAAAAGAGGACATCTTATGGGTGCCACTGAAATGAAAGGGT  720

Qy   2830 ATTTGGGTTCCCAGTCCACTATACTGACGTCTCCAACATGACGCGCTTGGCGTGAGGCAGAG 2889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATTTGGGTTCCCAGTCCACTATACTGACGTCTCCAACATGACGCGCTTGGCGTGAGGCAGAG  780

Qy   2890 ACTGCTGGGCCGGTCATCGGACGTGCCCAGTCATCCGCCCACTCTTCTCCCTCCGCGTGAAGGA 2949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ACTGCTGGGCCGGTCATCGGACGTGCCCAGTCATCCGCCCACTCTTCTCCCTCCGCGTGAAGGA  840

Qy   2950 GTATTTGCGTGTGTGTAAGGGACATGGGGGCAAACTGAGGTAGCGACACAAAGTTAAAC 3009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GTATTTGCGTGTGTGTAAGGGACATGGGGGCAAACTGAGGTAGCGACACAAAGTTAAAC  900

Qy   3010 AAACAAACAAAAAACACAAAAACATAATAAAAACACACAAGAAACATGAGGATGGACGAGAAGTA 3069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAACAAACAAAAAACACAAAAACATAATAAAAACACACAAGAAACATGAGGATGGACGAGAAGTA  960

Qy   3070 TCAGCACCCAGAAGAGAGAGAAAAAGGAATTTAAAAACAAAAACCACAGAGGGGGGAAATACCGGA 3129
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TCAGCACCCAGAAGAGAGAGAAAAAGGAATTTAAAAACAAAAACCACAGAGGGGGGAAATACCGGA 1020

Qy   3130 GGGCTTGCCTTTGGCCTTGGCAAAAGGGTTGGGACATCATCTCTCGATTTTCAATGTGTTATTCTTCA 3189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GGGCTTTGCCTTTGCCTTGGCAAAAGGGTTGGACATCATCTCTGATTTTCAATGTTATTCTTCA 1080

Qy   3190 GTCCTATTTAAAAACAAAACCAAGCTCCCCTTCCCTTCCCCTTCCCTTTTTTTTTTCGG 3249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GTCCTATTTAAAAACAAAACCAAGCTCCCCTTCCCTTCCCCTTCCCTTTTTTTTTTCGG 1140

Qy   3250 TCAGACCTTTTATTTTCTACTCTTTTCTGTTTGTTGTTTGTTTGTTGTTTG 3309
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TCAGACCTTTTATTTTCTACTCTTTTCTGTTTGTTGTTTGTTTGTTGTTTC 1200

Qy   3310 TTGCTGTGACTGAAACAAGCAAGAGGTTATTGCAGCAAAAATACAGTAACAAAAAAATAGTAACA 3369
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TTGCTGTGACTGAAACAAGCAAGAGGTTATTGCAGCAAAAAATCAGTAACAAAAAAATAGTAACA 1260

Qy   3370 ATACCTTGCAGGAGGAAGGTGGGAGGGAGAAATTTTTAAAAGAAATCTAT 3429
          |||||||||||||||||||||||||||||||| ||||||
Db   1261 ATACCTTGCAGAGGAGGAAAGGTG---GGAGAGGAGAAAAAAGGAAATTCTATAGAAAATCTAT 1317

Qy   3430 ATATTGGGTTGTTTTTTTTTTGTTTTGGGTTTTTTTTTTTTTTTTTACT 3489
```

```
Db   1318 ATATTGGGTTGTTTTTTTTTGTTGTCTCTAGCCTGATCAGATAGGAGCACAAGCACGAGGGGGACGGAA 1376

Qy   3490 ATATATCTTTTTTTTTGTTGTCTCTAGCCTGATCAGATAGGAGCACAAGCACGAGGGGGACGGAA 3549
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1377 ATATATCTTTTTTTTGTTGTCTCTAGCCTGATCAGATAGGAGCACAAGCACGAGGGGGACGGAA 1436

Qy   3550 AGAGAGAGACACTCAGGGGCGGCAGCATTCCCTCCCACCGAGCTCTGCGTGCCAGCACC 3609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1437 AGAGAGAGACACTCAGGGGCGGCAGCATTCCCTCCCACCGAGCTCTGCGTGCCAGCACC 1496

Qy   3610 ATTCCTGGTCAGCCAAAACAGAACCCAGTTAGCAGCGAGGAGGACGAGAACACCACACAG 3669
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1497 ATTCCTGGTCACGCAAAACAGAACCCAGTTAGCAGCGAGGAGGACGAGAACACCACACAAG 1556

Qy   3670 ACATTTTTCTACAGTATTTCAGGTGCCTACCACACAGGAAAACCTTGAGAAAATCAGTTT 3729
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1557 ACATTTTTCTACAGTATTTCAGGTGCCTACCACACAGGAAAACCTTGAGAAAATCAGTTT 1616

Qy   3730 CTAGAACCCGGCTGTTACCCTCTTGTTTACAGTTTATATATATGATAGATATGAGATATA 3789
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1617 CTAGAACCCGGCTGTTACCCTCTTGTTTACAGTTTATATATATGATAGATATGAGATATA 1676

Qy   3790 TATATAAAAAGGTACTGTTAACTACTACTGTTACAA 3820
          ||||||||||||||||||||||||||||||||| ||
Db   1677 TATATAAAAGGTACTGTTAACTACTACTGTTAAAA 1707
```

Center project name: H_NH0458N05

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping Information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional Information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E...
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-579I5; the clone sequenced
to the right is RP11-579I5. Actual start of this clone is at base
position 1 of RP11-458N5; actual end is at base position 19752 of
RP11-579I5.

Ambiguous base exists at 127455. Single plasmid region exists
between 126594 and 126656. Data from AC104699 was used to finish
AC012074.

FEATURES   Location/Qualifiers
 source   1..197652
       /organism="Homo sapiens"
       /db_xref=taxon:9606"
       /chromosome="2"
       /map="2"
       /clone="RP11-458N5"
       /clone_lib="RPCI-11"
 repeat_region 98..161
       /rpt_family="GA-rich"
 repeat_region 1121..1272
       /rpt_family="(TA)n"
 repeat_region 1326..1625
       /rpt_family="Alu"
 repeat_region 1664..1699
       /rpt_family="AT_rich"
 repeat_region 1992..2266
       /rpt_family="Alu"
 repeat_region 2287..2443
       /rpt_family="MIR"
 repeat_region 4014..4143
       /rpt_family="Alu"
 repeat_region 4144..4454
       /rpt_family="Alu"
 repeat_region 4455..4657
       /rpt_family="Alu"
 repeat_region 4695..4781
       /rpt_family="L1"
 repeat_region 4782..5086
       /rpt_family="Alu"
 repeat_region 5099..5336
       /rpt_family="Alu"

 repeat_region 5336..5589
       /rpt_family="L1"
 repeat_region 5815..5955
       /rpt_family="MIR"
 repeat_region 6009..6320
       /rpt_family="Alu"
 repeat_region 6574..6726
       /rpt_family="L2"
 repeat_region 6951..7139
       /rpt_family="MIR"
 repeat_region 7378..7481
       /rpt_family="MIR"
 repeat_region 7508..7689
       /rpt_family="L2"
 repeat_region 7701..7946
       /rpt_family="Alu"
 repeat_region 7947..8214
       /rpt_family="L2"
 repeat_region 8215..8515
       /rpt_family="Alu"
 repeat_region 8516..8538
       /rpt_family="L2"
 repeat_region 8539..8667
       /rpt_family="T2_type"
 repeat_region 8668..8868
       /rpt_family="L2"
 repeat_region 8869..9176
       /rpt_family="Alu"
 repeat_region 9177..9406
       /rpt_family="L2"
 repeat_region 9460..9534
       /rpt_family="L2"
 repeat_region 9660..9965
       /rpt_family="Alu"
 repeat_region 9998..10034
       /rpt_family="(TG)n"
 repeat_region 10651..10815
       /rpt_family="Alu"
 repeat_region 10816..10835
       /rpt_family="(TTG)n"
 repeat_region 10836..11121
       /rpt_family="Alu"
 repeat_region 11389..11683
       /rpt_family="Alu"
 repeat_region 12188..12529
       /rpt_family="Alu"
 repeat_region 12975..13131
       /rpt_family="L1"
 repeat_region 13167..13466
       /rpt_family="Alu"
 repeat_region 13470..13533
       /rpt_family="L1"
 repeat_region 13765..14072
       /rpt_family="Alu"
 repeat_region 14216..14235
       /rpt_family="(TTTA)n"
 repeat_region 14241..14519
       /rpt_family="Alu"
 repeat_region 14890..15178
       /rpt_family="CR1"
 repeat_region 15717..15841
       /rpt_family="L1"
 repeat_region 16187..16217
       /rpt_family="AT_rich"
 repeat_region 21273..21317
       /rpt_family="(A)n"
 repeat_region 22691..22992
       /rpt_family="Alu"
 repeat_region 23576..23872
       /rpt_family="Alu"
 repeat_region 24865..24966
       /rpt_family="L1"
 repeat_region 24994..25287
       /rpt_family="Alu"

repeat_region   26925..27112   /rpt_family="Alu"
repeat_region   28773..28875   /rpt_family="MER1_type"
repeat_region   30336..30532   /rpt_family="Alu"
repeat_region   30533..30823   /rpt_family="L1"
repeat_region   30824..30998   /rpt_family="L1"

Query Match   32.8%;   Score 1408.6;   DB 9;   Length 197652;
Best Local Similarity   99.1%;   Conservative 0;   Pred. No. 5.1e-278;
Matches 1438;   Mismatches 9;   Indels 4;   Gaps 2;

```
Qy   2825 AGGGTATTTGGTTTCCCAGTCCCACTATACTGAGCTCCCAACATGAGCCGCTTGGCGAGG 2884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21930 AGGGTATTTGGTTTCCCAGTCCCACTATACTGAGCTCCCAACATGAGCCGCTTGGCGAGG 21871

Qy   2885 CAGAGACTGCTGGGCGGCCGGGTCATGGAGCGGGTGCCAGTCACTCATCCCACCCTCCCTG 2944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21870 CAGAGACTGCTGGGCGGCCGGGTCATGGAGCGGGTGCCAGTCACTCATCCCACCCTCCCTG 21811

Qy   2945 AAGGAGTATTTTGCGTGTGTGTAAGGGGACATGGGGCAAAACTGAGGTAGCGACACAAAGT 3004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21810 AAGGAGTATTTTGCGTGTGTGTAAGGGGACATGGGGCAAAACTGAGGTAGCGACACAAAGT 21751

Qy   3005 TAAACAAAACAAAACAAAAACACAAAAACATAATAAAACACCAAGAACATGAGGATGGGAGAG 3064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21750 TAAACAAAACAAAACAAAAACACAAAAACATAATAAAACACCAAGAACATGAGGATGGGAGAG 21691

Qy   3065 AAGTATTCAGCACCCAGAGAAGAGAGAGAAAGAGAAATTTAAAAAACCACAGAGGCGGGAAATA 3124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21690 AAGTATTCAGCACCCAGAGAAGAGAGAGAAAGAGAAATTTAAAAAACCACAGAGGCGGGAAATA 21631

Qy   3125 CCGGAGGGGCGTTTGCCTTGGCGAAAAAGGGTTGGACAATCATCATCTCGATTTTTCAATGTTATT 3184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21630 CCGGAGGGGCTTTGCCTTGGCGACATCATCATCTCCTGATTTTTTTCAATGTTATT 21571

Qy   3185 CTTCAGTCCTATTTAAAAAACAAAACCAAGCTCCCTTCCCTTCCCCCCCTTCCCTTTTTT 3244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21570 CTTCAGTCCTATTTAAAAAACAAAACCAAGCTCCCTTCCCTTCCCCCCCTTCCCTTTTTT 21511

Qy   3245 TTTCGGTCAGACCTTTTAATTTTCTACTCCTTTTCGTTTTGTTTTGTTTTGGGTTTTTTT 3304
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21510 TTTCGGTCAGACCTTTTAATTTTCTACTCCTTTTCGTTTTGTTTTGTTTTGGGTTTTTTT 21451

Qy   3305 GTTTCTTGCTGTGACTGAAACAAGAAGGTTATTGCCAGCAAAAATCAGTAACAAAAAATAG 3364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21450 GTTTCTTGCTGTGACTGAAACAAGAAGGTTATTGCCAGCAAAAATCAGTAACAAAAAATAG 21391

Qy   3365 TAACAAATACCTTGCAGAGGAAAGGGTGGGAGGGAAAAAGGGAAATTTTTTTAAAGAAA 3424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21390 TAACAAATACCTTGCAGAGGAAAGGTG---GGAGAGAGGAAAAAGGAAATTCTATAGAAA 21334

Qy   3425 TCTATATAATTAGGGGTTGTTGTTTTTGTTTTTTTGTTTTTTGTTTTTTTTGGGGTTTTTTT 3484
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21333 TCTATATAATTAGGGGTTGTTGTTTTTGTTTTTTTGTTTTTTGTTTTTTTTGGG-TTTTTTT 21275

Qy   3485 TTACTATATATCTTTTTTTTTTTGTTGTCTTAGCCTGATCAGATAGGAGCACAAGCAGGGGA 3544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21274 TTACTATATATCTTTTTTTTTTTGTTGTCTTAGCCTGATCAGATAGGAGCACAAGCAGGGGA 21215

Qy   3545 CGGAAAGAGAGAGAGACACTCAGGCGCGGCCAGCCATTCCCTCCCCAGCCACTGAGCTGTGCCA 3604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21214 CGGAAAGAGAGAGACACTCAGGCGCGGCATTCCCTCCCCAGCCACTGAGCTGTCGTGCCA 21155

Qy   3605 GCACCATTCCTGGGTCACGCAAAAACAGAGAACCCAGTTAGCACCAGGGGAGCAGAACACCAC 3664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21154 GCACCATTCCTGGGTCACGCAAAAACAGAGAACCCAGTTAGCACCAGGGAGCAGAACACCAC 21095

Qy   3665 ACAAGACATTTTTCTACACAGTATTCAGGTGCCTACCACACAGGGAAAACCTTGAAGAAAATC 3724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21095 GCACCATTCCTGGTCACGCAAAAACAGAGAACCCAGTTAGCACCAGGGAGCAGAACACCAC 21035
```

```
Db  21094 ACAAGACACATTTTTCTACAGTATTCAGGTGCCTACCACACAGGGAAAACCTTGAAGAAAATC 21035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3725 AGTTTCTAGAAGCCGCTGTTACCCTCTTGTTTACAGTTATATATATATATGATAGTATGAG 3784

Db  21034 AGTTTCTAGAAGCCGCTGTTACCCTCTTGTTTACAGTTATATATATATATGATAGTATGAG 20975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3785 ATATATATATTAAAAAGTTACTGTTAAACTACTGTCATAACCCGACTTCATAATGGTGCTTTCA 3844

Db  20974 ATATATATATTAAAAAGTTACTGTTAAACTACTGTCATAACCCGACTTCATAATGGTGCTTTCA 20915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3845 AACAGCGAGATGAGTAAAAAACATCAGCTTCCACGTTGCCTTCTCGGCCAAAGGGTTTCACC 3904

Db  20914 AACAGCGAGATGAGTAAAAAACATCAGCTTCCACGTTGCCTTCTCGGCCAAAGGGTTTCACC 20855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3905 AAGGATGGGAGAAAGGGAGACACCTTGCAGATGGCGCGGTTCTCACGGTGGGCTTCCCCT 3964

Db  20854 AAGGATGGGAGAAAGGGAGACACCTTGCAGATGGCGCGGTTCTCACGGTGGGCTTCCCCT 20795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3965 TGGTTTGTTAACGAAGTGAAAGAGAGAGAACTTGGGAGCCAGGTTCTCCCTGCCAAAAGGG 4024

Db  20794 TGGTTTGTTAACGAAGTGAAAGAGAGAGAACTTGGGAGCCAGGTTCTCCCTGCCAAAAGGG 20735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4025 GGCTAGATGAGGTGGTTCGGGCCGTGACACAGCGTCGAGAGGTCGGGAATTCATCCAGACTCATGC 4084

Db  20734 GGCTAGATGAGGTGGTTCGGGCCGTGACAGCGTGGATTCATCCAGACTCAATGC 20675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4085 AATAACCCTTTGATTGTTTTCTAAAAAGGAGACTCCCTCGGCCAAGATGGCAGAGGGTACGG 4144

Db  20674 AATAACCCTTTGATTGTTTTCTAAAAAGGAGACTCCCTCGGCCAAGATGGCAGAGGGTACGG 20615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4145 AGTCTTCAGGCGCCCAGTTTCTCACTTTACCAATTCGAGGGCTCCTTGGTTGGGGGATCAGA 4204

Db  20614 AGTCTTCAGGCGCCCAGTTTCTCACTTTACCAATTCGAGGGCTCCTTGGTTGGGGGATCAGA 20555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4205 ACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAAACCCCACCTGGACAAATAAAAAAAACA 4264

Db  20554 ACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAAACCCCACCTGGACAAATAAAAAAAACA 20495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   4265 TACAAAAACGTA 4275

Db  20494 TACAAAAACGTA 20484
```

RESULT 8
AC120824
LOCUS   AC120824   123936 bp   DNA   linear   HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-282D16, *** SEQUENCING IN PROGRESS
            ***, 38 unordered pieces.
ACCESSION   AC120824
VERSION   AC120824.2 GI:21902541
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 123936)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C.C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

```
AUTHORS   Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
          Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
          Kratovic,J., Kureshi,N., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
          Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
          Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
          Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
          Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
          Moser,M., Neal,D., Newsom,J., Newton,N., Nguyen,A., Nguyen,N.,
          Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
          Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
          Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
          Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
          Scherer,S., Scott,G., Shen,H., Shoosntari,N., Sisson,I.,
          Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
          Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
          Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
          Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
          Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
          Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
          Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
          Weinstock,G. and Gibbs,R.
TITLE     Direct Submission
JOURNAL   Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 2  (bases 1 to 123936)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 3  (bases 1 to 123936)
TITLE     Direct Submission
JOURNAL   Submitted (13-MAY-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jul 18, 2002 this sequence version replaced gi:20531783.
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -------- Project Information --------
          Center project name: GWLV
          Center clone name: CH230-282D16
          -------- Summary Statistics --------
          Sequencing vector: Plasmid;
          Chemistry: Dye-terminator Big Dye: 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 95907 bases at least Q40
          Consensus quality: 99223 bases at least Q30
          Consensus quality: 101036 bases at least Q20
          ---------------------------------------
          NOTE: Estimated insert size may differ from sequence length
              (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          --------
          NOTE: This is a 'working draft' sequence. It currently
          consists of 38 contigs. The true order of the pieces
          is not known and their order in this sequence record is
          arbitrary. Gaps between the contigs are represented as
          runs of N, but the exact sizes of the gaps are unknown.
          This record will be updated with the finished sequence
          as soon as it is available and the accession number will
          be preserved.
          --------
                     1     1009: contig of 1009 bp in length
                  1010     1109: gap of unknown length
                  1110     2113: contig of 1004 bp in length
                  2114     2213: gap of unknown length
                  2214     3233: contig of 1020 bp in length
                  3234     3333: gap of unknown length
                  3334     4442: contig of 1109 bp in length
                  4443     4542: gap of unknown length
                  4543     6202: contig of 1660 bp in length
                  6203     6302: gap of unknown length
                  6303     7464: contig of 1162 bp in length
                  7465     7564: gap of unknown length
                  7565     8772: contig of 1208 bp in length
                  8773     8872: gap of unknown length
                  8873    10371: contig of 1499 bp in length
                 10372    10471: gap of unknown length
                 10472    11714: contig of 1243 bp in length
                 11715    11814: gap of unknown length
                 11815    12924: contig of 1110 bp in length
                 12925    13024: gap of unknown length
                 13025    14347: contig of 1323 bp in length
                 14348    14447: gap of unknown length
                 14448    15652: contig of 1205 bp in length
                 15653    15752: gap of unknown length
                 15753    17796: contig of 2044 bp in length
                 17797    17896: gap of unknown length
                 17897    18937: contig of 1041 bp in length
                 18938    19037: gap of unknown length
                 19038    21543: contig of 2506 bp in length
                 21544    21643: gap of unknown length
                 21644    24540: contig of 2897 bp in length
                 24541    24640: gap of unknown length
                 24641    28112: contig of 3472 bp in length
                 28113    28212: gap of unknown length
                 28213    31101: contig of 2889 bp in length
                 31102    31201: gap of unknown length
                 31202    34415: contig of 3214 bp in length
                 34416    34515: gap of unknown length
                 34516    37353: contig of 2838 bp in length
                 37354    37453: gap of unknown length
                 37454    40938: contig of 3485 bp in length
                 40939    41038: gap of unknown length
                 41039    44331: contig of 3293 bp in length
                 44332    44431: gap of unknown length
                 44432    48035: contig of 3604 bp in length
                 48036    48135: gap of unknown length
                 48136    51181: contig of 3046 bp in length
                 51182    51281: gap of unknown length
                 51282    55361: contig of 4080 bp in length
                 55362    55461: gap of unknown length
                 55462    59580: contig of 4119 bp in length
                 59581    59680: gap of unknown length
                 59681    63525: contig of 3845 bp in length
                 63526    63625: gap of unknown length
                 63626    67434: contig of 3809 bp in length
                 67435    67534: gap of unknown length
                 67535    72194: contig of 4660 bp in length
                 72195    72294: gap of unknown length
                 72295    77677: contig of 5383 bp in length
                 77678    77777: gap of unknown length
                 77778    81349: contig of 3572 bp in length
                 81350    81449: gap of unknown length
                 81450    86396: contig of 4947 bp in length
                 86397    86496: gap of unknown length
                 86497    91224: contig of 4728 bp in length
                 91225    91324: gap of unknown length
                 91325    97769: contig of 6445 bp in length
                 97770    97869: gap of unknown length
                 97870   105670: contig of 7801 bp in length
                105671   105770: gap of unknown length
                105771   111368: contig of 5598 bp in length
                111369   111468: gap of unknown length
                111469   116680: contig of 5212 bp in length
                116681   116780: gap of unknown length
                116781   123936: contig of 7356 bp in length.
FEATURES             Location/Qualifiers
     source          1..123936
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-282D16"
```

BASE COUNT   35904 a   21413 c   21468 g   36486 t   8665 others
ORIGIN

Query Match   28.5%;   Score 1221.4;   DB 2;   Length 123936;
Best Local Similarity   81.8%; Pred. No. 1.2e-239;
Matches 1416;   Conservative   0;   Mismatches 286;   Indels 33;   Gaps 5;

```
Qy    868 GGCTGAGAGAAGAAAACCAGGGCGC 927
          ||||||||||||||||||||||||||
Db  84794 GGTTGAGAAGAAAAGCCAAGGTGATTGCAGTAATCAATGCTGTGTGGAGGAAAACCAGGGCTC 84853

Qy    928 CGGGGAGTCTCAGAAGGTGGAGGAGGGCCAGCCCCACTGACCC 987
          |||||||||||||||||||||||
Db  84854 TGGGGAGTCTCAGAGAGGTGGAGCAGGCCACCAGCCCCA--GGACCC 84912

Qy    988 CGCATCCCCACTGTGTGGCTACCACGCGCCTGAGCCCGTGGGGTCCGAGGACAAGAA 1047
          |||||||||||||||||||||||||||||||||||
Db  84913 TGCACTTCCACTGTGGCCATTACCCCTGTGCCAGTAGGGGCTGATGCTGGGGACAAGAA 84972

Qy   1048 TGCCACCAAAGCAGGCGGATGACGAGCCAGAGTAGGACGGGCGGGCGTTGGCATTGG 1107
          |||||||||||||||||||||||
Db  84973 TGCTATCAAGGCAGCTGATGAGGACCCCGAGTATGAGGAATGGCCCAGGGGCCTTGGGCATTGG 85032

Qy   1108 GGAGCTGGGTGTGGGGGAAAACGCGGGGGCCAGCCCCGGACCCAGTTGTGTCTTTG 1167
          |||||||||||||||||||||||
Db  85033 AGAGCTGGTGTGGGGGGAAACTCCGGGGGCTTCTCTGGTGGGCCAAGCCGAATTGTTGTTGTCTGTTG 85092

Qy   1168 GTGGAATGACGGGCCCGGAGCCAGCTGAAGGCACCCGGGGTCATGTGGTTCGGAGA 1227
          |||||||||||||||||||||||
Db  85093 GTGGGATGACAGGCCGAGAGCGAAGGCACTCGGGTCAGTGGTTTGGAGA 85152

Qy   1228 CGGCAAAATTCTCAGTGGTGTGTGTTGAGAAGCGATGCCGGCTCGTTTTGCAGTGC 1287
          |||||||||||||||||||||||
Db  85153 TGGGCAAAGTTTTCAGTGCATTGTGTCGAGAAGCTCATTGCCCCTGAGCCTCCTTCGCA---- 85208

Qy   1288 GTTCCACCAGGCCACGTACAACAAGCAGCCACCCATAGTACCGCAAAGCCATCTACGGAGGTGCT 1347
          ||||||||||||||||||||||| |
Db  85209 TTTCCACCAGGCCACCTACAACAAGCAGCCCCATGTACCCGCAAAGCCATCTACGGAAGTGCT 85268

Qy   1406 AGTGACACTGCCAAGGCCGTGGAGGTGCAGAAACAGCCCATGATTGAAATGGGGCCCTGGGG 1465
          ||||||||||||||||||||||||
Db  85269 CCAGGTGGCCAGCAGCTGTTCAGGGAACCTGTTCCCAGCCGTGTCATGACACGGACAGTGAA 85328

Qy   1466 GGCTTCCAGCCTTCTGCCCTAAGGGCCTGGAGCCACCCAGAAGAAGAGAAGAATCCCTAC 1525
          |||||||||||||||||||||||
Db  85329 AGTGACACTGGCCAAGGCTGTGGAGGTGCAGAAACAGCCCATGATTGAGTGGGGCCCTGGGG 85388

Qy   1526 AAAGAAGTGTACACGGACGACATGTGGGGTGGAACCTGAGGCAGCTGCCTACGCACCACCTCCA 1585
          ||||||||||||||
Db  85389 GGCTTCCAGCCCTCTGGTCCCAAAGGGCCTGGGAGCGCCACCACCACAGAGAACAAATCTTTTAC 85448

Qy   1586 CCAGCCAAAAAAGCCCCGGAAGAGCACAGCGGGAGAAGCCCAAGGTCAAGGAGATTATTGATT 1645
          |||||||||||||||||| |
Db  85449 AAGGAAGTTTACACCGGACATGTGGGCTGGAGCCTGAGCGCAGCTGCTTATTCCCCACCCCCA 85508

Qy   1646 GAGCGCCACAAGAGAGGCGGGCTGGTTGGTTGCTGCGACAAGTGCGACAGCCATTGAGGAC 1705
          ||||||||||||||||
Db  85509 CCAGCCAAGAAGAAACCCAGAAAGAGACCAC------AACCTAAGGTCAAGGAGGATCATTGAT 85560

Qy   1706 ATCTGCATCTTCCTGTGTCGGGAGCCTCAATGTTACCCTGGACACACCCCCCTCTTCCGTTGGAGGA 1765
          ||||||||||||||||||
Db  85561 GAACACACAAGAGAGAACGACTCATGTATGAGGTGCGCCCAGAGTGCCCAAAACATCGAGGAC 85620

Qy   1766 ATGTGCCAAAAACTGCAAGAGAACTGCCTTCTTTCTGGAGGTGCGGTACCAGTACGACAGCGACGGC 1825
          |||||||||||||
Db  85681 ATGTGCCAGAACTGTAAGAACTGCTTCTTTGGAGTGTGCTTACCAATACGATGACGATGTG 85740

Qy   1826 TACCAGTCCTACTGCACCACTCTGCTCTGTGGGGGAAACAAC 1885
          |||||||| ||| |||||||||||||||||||
Db        TACCAGTCCTTCTGTACCACTCTGCTGGGGGAAACAAC 86497
```

(second column)

```
Db  85741 TACCAGTCCTTCTGTACCACTCTGCTCTGTGTTGGGGGCGGCGGGAGAGTGGTCATGTGTGGGACAAC 85800
          |||||||||||||||||||||
Qy   1886 AACTGGCTGCAGGTGCTGTTTTGCGTGGTGAGTGTGGACCTCGTGGAGTGGCTGGGGGGCTGGGGTGGGTGGGGCGGGGCTGCC 1945
          ||| |||||||||
Db  85801 AACCGGCTGCAGGTGCTGTTTTGCGTGGTCAGTGTGGAGTGTGTTGGTCTGTTGGTGTGGGGGGCCAGGGGTGCC 85860

Qy   1946 CAGGCCAGCCAATTAAGGAGAGACCCCTGGAACTGCTACATGTGCGGGGCCACAAGGGTACCTAC 2005
          ||||||||||||||||||||||
Db  85861 CAAGCAGCCATTAAGGAGAGACACCCCTGGAAACTGCTACACATGTG---------------T 85902

Qy   2006 GGGGCTGCTGCGGCGGGCGAGAGAGGGACTGCGCCCTCCCGGCGTCCAGGAATGTTCTTCGCTAATAAC 2065
          ||||||||||||||||||
Db  85903 GAGCTGCTGCGGGAGGAGGAGGGACTGGCCTTCTCTGGACTGCCAGATGTTTCTTTTGCCAATAAC 85962

Qy   2066 CACGACCAGGAATTTGACCCCTCCAAAGGTTTACCCACCTGTCCCACCTGTGAGAAGAGGAAG 2125
          |||||||||||||||||||||||
Db  85963 CACGACCAGGAATTTGACCCCCCGAAGGTTTAACCCCACCCTGTCCCAGCTGAGAAGAGGAAG 86022

Qy   2126 CCCAATCCGGGTGCTGGTCTCTCCTTTGGATGGAAATGGCATTGGGTGCCATTGG 2185
          |||||||||||||||||||||||
Db  86023 CCCAACCCAGGTGGCTTACCTCTATTTGATGGAAATGGCCTTGGGTGTGCGGGCTGAAGGAC 86082

Qy   2186 TTGGGCATTCAGGTGGACACCGGCTACATTGCCTCGGAGGTGTGTGAGGACTCCATCAGGGTG 2245
          |||||||||  | |||||
Db  86083 CTGGGCATCCAAAGTGGAACCGGCTACAATTGCCTTCCGAGGGTGTGCGAGGGACTCCATCACCGGTG 86142

Qy   2246 GGCATGGTGCGGGGCACCAGGGGAAGAGATCATCATGTACGTCGGGGGACGTCCCGCACGCTCACACAG 2305
          ||||||||||||||||||
Db  86143 GGCATAGTGCGGCACCACAGGGAAAGATCAATGTTCGTCGGGGACGTCTCGCAGCATCACACAG 86202

Qy   2306 AAGCATATCCAGGAGTGCGGGGCCCATTCCGATCTGTGATTCGGTGGATGTCCCTGCAAATGAC 2365
          ||||||||||||||||||
Db  86203 AAGCATATCCAGGAGAGGGGGCCCATACACAATCCTGGATTTGGGGACCAGTCCTTGCAAATGAC 86262

Qy   2426 GAGTTCTACCCGCCTCCTGCATGATGCGCGGCCCAAGAGGAGGAGAAGATGAATCGCGCCCTTCTTC 2485
          ||||||||||||||
Db  86263 CTCTCCCATCATCAACCCTGCCCGCAAAGGGACTTTACAGGGCACGGCCAGCCGCTTCTTTT 86322

Qy   2486 TGGCTCTTTGAGAATGTGGGTGGCGGCGTTAGTGACAAGAGAGGACATCTCGCGAATTT 2545
          ||||||||||||
Db  86323 GAGTTCTACACCCTCCTGCATGAGTGCGGGCCCAAGAGGAGGAGATCGCCCCTTCTTC 86382

Qy   2546 CTCGAGTCCAACCCTGTGATGAATTGATCGCCAAAGAAGTGTCAGCTGCCACACAGGG 2600
          |
Db  86383 TGGCTCTTTGAGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 86442

Db  86443 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 86497
```

RESULT 9
AC009474

```
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (24-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 168651)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (27-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 168651)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jun 27, 2000 this sequence version replaced gi:7630880.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: saplens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0179G23
            --------------
NOTICE:     This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
VECTOR:     pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-547F18; the clone sequenced
            to the right is RP11-444B4, 200 base pair overlap.  Actual end of
            this clone is at base position 18711 of RP11-444B4.
FEATURES             Location/Qualifiers
     source          1..168651
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-179G23"
                     /clone_lib="RPCI-11"
     repeat_region   13..497
                     /rpt_family="L2"
     repeat_region   616..683
                     /rpt_family="L2"
     repeat_region   1615..1815
                     /rpt_family="MIR"
```

```
     repeat_region   2577..2748
                     /rpt_family="MIR"
     repeat_region   2954..3144
                     /rpt_family="MER1_type"
     repeat_region   3149..3465
                     /rpt_family="L1"
     repeat_region   3510..3567
                     /rpt_family="L1"
     repeat_region   3569..3760
                     /rpt_family="MER1_type"
     repeat_region   3827..4198
                     /rpt_family="MER1_type"
     repeat_region   5656..6129
                     /rpt_family="MaLR"
     repeat_region   6205..6316
                     /rpt_family="Alu"
     repeat_region   6428..6686
                     /rpt_family="L2"
     repeat_region   7247..7437
                     /rpt_family="L1"
     repeat_region   7819..7868
                     /rpt_family="MER53"
     repeat_region   9304..9469
                     /rpt_family="CR1"
     repeat_region   10317..10439
                     /rpt_family="Alu"
     repeat_region   12219..12455
                     /rpt_family="MIR"
     repeat_region   12467..13337
                     /rpt_family="Alu"
     repeat_region   13553..14453
                     /rpt_family="L1"
     repeat_region   14853..15123
                     /rpt_family="L1"
     misc_feature    16209..16345
                     /note="similar to EST AA338793 (NID:g1991103)"
     misc_feature    16209..16345
                     /note="similar to EST AI123298 (NID:g3539064) ov48c08.x1"
     misc_feature    16210..16345
                     /note="similar to EST AL037015 (NID:g5928008)"
     repeat_region   16513..16676
                     /rpt_family="ERVL"
     repeat_region   16970..17198
                     /rpt_family="L1"
     repeat_region   17199..17548
                     /rpt_family="MER1_type"
     repeat_region   17575..18261
                     /rpt_family="L1"
     repeat_region   19409..19531
                     /rpt_family="L1"
     repeat_region   19540..19659
                     /rpt_family="Alu"
     misc_feature    25486..25822
                     /note="similar to EST AI928247 (NID:g5664211) wo95h11.x1"
     repeat_region   26074..26366
                     /rpt_family="Alu"
     repeat_region   26384..26542
                     /rpt_family="MIR"
     repeat_region   26569..27058
                     /rpt_family="L2"
     repeat_region   27140..27647
                     /rpt_family="L2"
     repeat_region   30824..31150
                     /rpt_family="MaLR"
     repeat_region   32262..32525
                     /rpt_family="CR1"
     repeat_region   32568..32770
                     /rpt_family="MER1_type"
     repeat_region   32775..33032
                     /rpt_family="CR1"
     repeat_region   33172..33418
                     /rpt_family="MIR"
     repeat_region   34930..35230
                     /rpt_family="MIR"
```

repeat_region /rpt_family="Alu"   35302. .35555
repeat_region /rpt_family="L2"   35975. .36081
repeat_region /rpt_family="L2"   36364. .36515
repeat_region /rpt_family="Alu"   37266. .37340
repeat_region /rpt_family="L2"   37400. .37440
repeat_region /rpt_family="L2"   37615. .37715
repeat_region /rpt_family="MIR"   38006. .38177
repeat_region /rpt_family="MER1_type"   43171. .43204
repeat_region /rpt_family="U2"   44777. .44926
repeat_region /rpt_family="MIR"   45905. .46110
repeat_region /rpt_family="MIR"   48479. .48638
repeat_region /rpt_family="MER1_type"   48914. .49033
repeat_region /rpt_family="MIR"   49069. .49337
repeat_region /rpt_family="MER1_type"   49340. .49621
repeat_region /rpt_family="Alu"   50377. .50498
repeat_region /rpt_family="MIR"   56393. .56545
repeat_region /rpt_family="MIR"   58109. .58420

Query Match 27.0%; Score 1157.6; DB 9; Length 168651;
Best Local Similarity 70.7%; Pred. No. 1.6e-226;
Matches 2010; Conservative 0; Mismatches 449; Indels 383; Gaps 20;

Qy   834 AGCGGGACGAGTGGCTGGCACGCCTGGGAAAAAGGGAGGGCTGAGAAGAAA-GCCAAGGTCATT   892
           |||||||||||||||||||||||||||||||||||||||||| ||| |||| || ||||||||
Db 75830 TCCGGAGGTTACAGTGAGCCGAGATAGCACCATTGCACTCCGGTGGTGACGAGAGAAGA 75889

Qy   1626 --------------------   1625
Db 75770 TGTGCCTGTAATCCCAGCTACTGGGGAGGCTGAGGCAGGAGAATCACTTGAACCAGGGAG 75829

[alignment block — rotated dense nucleotide sequence data]

```
Qy    2129  ATCCGGGTGCTGCTGTCTCTTCTTTTGATGGAATCGCTACAGGGCTCCTGGTTGCTGCTGAAGGACTTG  2188
            ||||  |||||||||||||| ||||||| || ||||  || |  |||||||||| ||||||  |||||||
Db   76425  ATCCAGGTGCTGCTGTCTTCTTTTGACGGGAATTGAATATGAGGGCTCCTCGGTTGCTGAATGGCTTG   76484

Qy    2189  GGCATTCAGGTGGACCGGCTACAGTTGCCCTGGGAGGTGTGTGAGGACTCCATCACGGTGGGC        2248
            |||||| ||| |||| | ||||| |||||| |||||||||||||||||| |||||| ||||||
Db   76485  GGCATTAGGTAGATACACTACA-TGTCTCAGGAGGTGTGTGAGGACGCCATCATTGTGGGC          76543

Qy    2249  ATGGTGCGGCACCAGGGAAGAGATCATGTACGTCGGGAGGTCCCAGCGTCACACAGAAG            2308
            ||| || |||||||||||  |||||| |||| | |||| |||||  ||||||||| |||||
Db   76544  ATGCTGTGGCACCAGGGAGGATCATGCACTCAGGGATTACCCGAACCATCATACACGAAG          76603

Qy    2309  CATATCCAGGAGTGGGGCCCCATTCGAATCTGGTGATTGGGGGGCAGTCCCTGCAAATGACCTC        2368
            || | |||||||||||||| |||||                   ||||||||||||||||||||
Db   76604  CACAACCAGGAGTGGGGCCCATTCT-------------------CTGCAATGACCTC              76641

Qy    2369  TCCATCGTCAACCTGCTCGCTCGGCAAAGGGCCTTACCAGGGCCACTGGGCCTCTTCTTTGAG        2428
            |||||  ||||| || | | |  || |||||  |||||  |||||| |||||||||||||||||
Db   76642  TCCATTGTCAACCCTGCCCACGAAGCACCTACAAAGGGACCACTGGGCCAGCAATTCTTTGAG        76701

Qy    2429  TTCTACCGCCTTCCTGCCGGATGATNGCGGGCCCCAAGGAGAGGGGAGATGAATCGCCCCTTCTTCTGG   2488
            ||| | || | | || | |||  || ||   | ||||||| |||| |||| |  | |||||||| | |||
Db   76702  TTTTCTCACCTT-GTGCATCARTGCACAGCCCAAGGAGGGAGAATGATC-ACCCTTCTTGTGG          76759

Qy    2489  CTCTTTGAGAAATGTGGTGGCCATGGGGGTTAGTGACAAGAGGGACACATCTCCGCGATTTCTC        2548
            ||  || ||| ||||  |||  ||  || ||||  |  |||| | ||||||||  |||| ||||||
Db   76760  CTCCTTGAGTGTGGTGGGCCATGGCCATTAGTCAAAAGAGGAGACATCTCACAATTTTTT          76819

Qy    2549  GAGTCCAACCCTGTGATTGATTGATCCCAAAGAAGTGTCAGCTGCACACAGGGCCCGCTAC          2608
            ||| | ||||  |||  ||| ||  |||  |||||||  || || | || |||  ||| ||||
Db   76820  TATTCGAATTCCATGATGAATTGCCAAGGAAGTGTCAACTGCACCCAGAACTACTAC          76879

Qy    2609  TTCTGGGGTAACCTTCCCGGTGTATGAACAGCGCCCTTGGCAATCCCACTGTGAAATGATAAGCTG    2668
            || | | |||| | ||  |||| | |||||| || || ||||  ||| | |||  |||| |||| |||
Db   76880  ATCTGTGGTAACCTTCCCAGTGAAAGAGGGCCATTGGCATTCACTATGAAATG-TAAGCTG         76938

Qy    2669  GAGCTGCAGGAGTGTCTGGACCATGGCCAGGATAGCCCAGTTCAGCAAAGTGAGGACCATT       2728
            |||||| | ||||| | |||||| ||||||| | | || ||| || || ||  || ||||||
Db   76939  GAGCTGCAGGAGTGTGTGGACCATGGCCAGAGAAAGGCCAAGTTCAGGAAAACGGAGGACCATT     76998

Qy    2729  ACTACGAGGTCAAAACTCCATAAAAGCAGAGGGCAAGAAAGACCAGCCAGGAGCAGCCGGCTAC    2788
            ||| |||| | ||| || | || || || || || ||||| ||| | |||  | |  | | |||
Db   76999  ACTACCAGGTCAAACTCCATAATGAACGGGGCAAAAAGACCAGGTGTCCCCGTCTTCATGAAT    77058

Qy    2789  GAGAAAGAGGACATCTTATGGTGCACTGAAATGGAAAAGGGTGTATTTGGTTTTCCCAGTCCAC     2848
            ||| | ||||| | |   || |||| ||||||| || ||||| |  |||| || |||||  |||
Db   77059  GAGAAGGGA----CACCTTAGGTGCATTGAAATTGAAAAGGGTATTTGGACTCCCGTGTCCAT     77114

Qy    2849  TATACTGACGTGTCCAACATGAGCCCGTTGGCGGAGCAGAACTGCCTGGGCCGGTCATGG      2908
            |||||| ||| |||| ||||| |||  | ||||| ||||||||| ||||||| |||||||
Db   77115  TATACTGACATCTCCAACATAAAGCCATTGGGCAAGGCCAGAGACGTGTTGCGGTGGACTCATGG     77174

Qy    2909  AGGGTGCCAGTCATCCGCCACCTCTTCTGCTCCGCTCAAAGGAGAGTATTTTGGCGTGTGTGTAA   2968
            ||  | | | ||||  || || || ||||  |  |||||||| |||||  |  | ||| ||||  |
Db   77175  AGCATAGCAGGTCACTGCCAATTCTCCCTCCAAGGGGTATTTTTGCTTTTGTGGTTGTAA      77234

Qy    3021  AAACACAAAACATAATAAAAACACCAAGACAACATGAGGAGAGAGAGAGAGTATCAGCACCCAG   3080
            || || |||||| | || | ||| |||| || || ||| || || |  |||  ||| | |||||
Db   77235  GGGACAATGGGGCGCAAACTGAGGTAGTGACACAAAGTTAAACACTATAAAAAAAAAAAAA     77294

Qy    2969  GGGACAATGGGGCGCAAACTGAGGTAGTGACACAAAGTTAAACAAAC----------AAAACAAA  3020
            |||  ||  ||  | |   |||| | | ||  ||||  |||||  ||           ||||| ||
Db   77059  GAGAAGGGA----CACCTTAGGTGCATTGAAATTGAAAAGGGTATTTGGACTCCCGTGTCCAT     77114

Qy    3081  AAGAG-AAAAAGGAATTTAAAACAAAAAACCACAGAGAGGCGAAATACCGGAGGGGCTTTGCC   3139
            |||| |  || || ||  | ||| |  |  | |  || || || |   | |  | ||||||||
Db   77295  CAAAACAAAACATAATAAAAACACTAACAATGTAAGAAAGAGAAGTATCAGCACCCAG      77354

Qy    3140  TTGGCAAAAAGGGTTGGACATCATCCTCTGATTTTTGATTTTCAAATGTTATTCTTTCAGTCTATTTA  3199
            ||||  |  ||||||||||  |  | |  |||||| | ||| | |||||  ||| |||| || |||||
Db   77415  TTGCAAAAAGGGTTGGA---CATCTTCTGATTTTTGAATGTGTTTTTCTTCAGTTCTATTTA        77471

Qy    3200  AAAACAAAACCAAGCTCCCTTCCCTTCCCCTCCCCCCCTCCCTTTTTTTGGGTCAGAGCCTTT   3259
            ||||||||| ||||||||||||| | ||||||||  ||||| | |||||| | |||||  || ||
Db   77472  AAAACAAAA-CAAGCTCCCTTCCTCTTCCCTCCCCGGCCCCCTCGCCCTCCCGGCCTTTTTGTTTCTTTT   77530

Qy    3260  TATTTTCTACTCTTTTTCAGAGGGGTTTTCTGTTTGGGTTTTGTTTCTTGTGTGAC    3319
            |||||||| |  || |||  | |  ||  | ||    ||||||||||||| | ||  |
Db   77531  CTTTTTTTTTTTTTTTTTTTTTTTGGTAAGAACTTTTGTTTCTTACTATTTCCATCTATGTT   77590

Qy    3320  TGAAACAAGAAAGGTTATTTGCAG  3341
            ||| | ||| |  || ||| ||||
Db   77591  GCTGCAAAGGACATTGCATGCAG   77612
```

RESULT 10

```
AC112586/c
LOCUS       AC112586             176697 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-28I22, *** SEQUENCING IN PROGRESS
            ***, 65 unordered pieces.
ACCESSION   AC112586
VERSION     AC112586.3  GI:21738467
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 176697)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Denn,A.L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Gardner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,U., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,C., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176697)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
```

```
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
REFERENCE  3  (bases 1 to 176697)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Jul 12, 2002 this sequence version replaced gi:20303171.
           --------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
           --------------- Project Information
              Center project name: GRAT
              Center clone name: CH230-28122
           --------------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 121959 bases at least Q40
              Consensus quality: 126371 bases at least Q30
              Consensus quality: 130428 bases at least Q20
           ---------------

           NOTE: Estimated insert size may differ from sequence length
              (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           NOTE: This is a 'working draft' sequence. It currently
              consists of 65 contigs. The true order of the pieces
              is not known and their order in this sequence record is
              arbitrary. Gaps between the contigs are represented as
              runs of N, but the exact sizes of the gaps are unknown.
              This record will be updated with the finished sequence
              as soon as it is available and the accession number will
              be preserved.

                   1    1415:  contig of 1415 bp in length
                1416    1515:  gap of unknown length
                1516    2955:  contig of 1450 bp in length
                2956    3065:  gap of unknown length
                3066    4228:  contig of 1163 bp in length
                4229    4328:  gap of unknown length
                4329    5385:  contig of 1057 bp in length
                5386    5485:  gap of unknown length
                5486    6670:  contig of 1185 bp in length
                6671    6770:  gap of unknown length
                6771    7813:  contig of 1043 bp in length
                7814    7913:  gap of unknown length
                7914    9673:  contig of 1760 bp in length
                9674    9773:  gap of unknown length
                9774   11470:  contig of 1697 bp in length
               11471   11570:  gap of unknown length
               11571   13162:  contig of 1592 bp in length
               13163   13262:  gap of unknown length
               13263   14504:  contig of 1242 bp in length
               14505   14604:  gap of unknown length
               14605   15974:  contig of 1370 bp in length
               15975   16074:  gap of unknown length
               16075   17396:  contig of 1322 bp in length
               17397   17496:  gap of unknown length
               17497   19323:  contig of 1827 bp in length
               19324   19423:  gap of unknown length
               19424   21041:  contig of 1618 bp in length
               21042   21141:  gap of unknown length
               21142   22419:  contig of 1278 bp in length
               22420   22519:  gap of unknown length
               22520   23942:  contig of 1423 bp in length
               23943   24042:  gap of unknown length
               24043   25333:  contig of 1291 bp in length
               25334   25433:  gap of unknown length
               25434   26530:  contig of 1097 bp in length
               26531   26630:  gap of unknown length
               26631   28270:  contig of 1640 bp in length
               28271   28370:  gap of unknown length
               28371   29970:  contig of 1600 bp in length
               29971   30070:  gap of unknown length
               30071   32265:  contig of 2195 bp in length
               32266   32365:  gap of unknown length
               32366   34179:  contig of 1814 bp in length
               34180   34279:  gap of unknown length
               34280   36392:  contig of 2113 bp in length
               36393   36492:  gap of unknown length
               36493   38552:  contig of 2060 bp in length
               38553   38652:  gap of unknown length
               38653   40932:  contig of 2280 bp in length
               40933   41032:  gap of unknown length
               41033   43021:  contig of 1989 bp in length
               43022   43121:  gap of unknown length
               43122   44977:  contig of 1856 bp in length
               44978   45077:  gap of unknown length
               45078   47358:  contig of 2281 bp in length
               47359   47458:  gap of unknown length
               47459   49554:  contig of 2096 bp in length
               49555   49654:  gap of unknown length
               49655   51876:  contig of 2222 bp in length
               51877   51976:  gap of unknown length
               51977   53794:  contig of 1818 bp in length
               53795   53894:  gap of unknown length
               53895   55634:  contig of 1740 bp in length
               55635   55734:  gap of unknown length
               55735   57445:  contig of 1711 bp in length
               57446   57545:  gap of unknown length
               57546   60210:  contig of 2665 bp in length
               60211   60310:  gap of unknown length
               60311   62174:  contig of 1864 bp in length
               62175   62274:  gap of unknown length
               62275   64285:  contig of 2011 bp in length
               64286   64385:  gap of unknown length
               64386   66764:  contig of 2379 bp in length
               66765   66864:  gap of unknown length
               66865   69398:  contig of 2534 bp in length
               69399   69498:  gap of unknown length
               69499   71451:  contig of 1953 bp in length
               71452   71551:  gap of unknown length
               71552   74405:  contig of 2854 bp in length
               74406   74505:  gap of unknown length
               74506   76586:  contig of 2081 bp in length
               76587   76686:  gap of unknown length
               76687   80397:  contig of 3711 bp in length
               80398   80497:  gap of unknown length
               80498   82299:  contig of 1802 bp in length
               82300   82399:  gap of unknown length
               82400   84796:  contig of 2397 bp in length
               84797   84896:  gap of unknown length
               84897   86981:  contig of 2085 bp in length
               86982   87081:  gap of unknown length
               87082   89794:  contig of 2713 bp in length
               89795   89894:  gap of unknown length
               89895   91696:  contig of 1802 bp in length
               91697   91796:  gap of unknown length
               91797   94458:  contig of 2662 bp in length
               94459   94558:  gap of unknown length
               94559   98352:  contig of 3794 bp in length
               98353   98452:  gap of unknown length
               98453  101134:  contig of 2682 bp in length
              101135  101234:  gap of unknown length
              101235  104588:  contig of 3354 bp in length
              104589  104688:  gap of unknown length
              104689  107761:  contig of 3073 bp in length
              107762  107861:  gap of unknown length
              107862  111340:  contig of 3479 bp in length
              111341  111440:  gap of unknown length
              111441  115374:  contig of 3934 bp in length
```

```
Query Match           25.1%;   Score 1078.4;   DB 2;   Length 176697;
Best Local Similarity 79.1%;   Pred. No. 2.9e-210;
Matches 1259;   Conservative   0;   Mismatches 306;   Indels 27;   Gaps 3;
```

```
Qy       1269 TGAGCTCGTTTTGCAGTGCGTTCCACCAGGCCACGTACAACAAGCAGCCCATGTACCGCA 1328
              |||| || ||  | |          | | |  | |||    |
Db      24084 TGAGTTCTTTAGGAATAAAACAATTCTACGTGAAGTAATTTANNNNNNNNNNNNNNNNNN 24025

Qy       1329 AAGCCATCTACGAGGTCCTGCAGGTGGCCAGCAGCCGCGCGGGGAAGCTGTTCCCGGTGT 1388

Db      24024 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 23965

Qy       1389 GCCACGACAGCGATGAGAGTGACACTGCCAAGGCCGTGGAGGTGCAGAACAAGCCCATGA 1448
                                    | ||  ||| |  ||||||||||| ||| |   |   |
Db      23964 NNNNNNNNNNNNNNNNNNNNNNCTCTCGCAACGGTGTGGAGGTGCATAACTATTGC-CAA 23906

Qy       1449 TTGAATGGGCCCTGGGGGGCTTCCAGCCTTCTGGCCCTAAGGGCCTGGAGCCACCAGAAG 1508
              |||| |||| ||| || ||||||||||| ||||| || ||||||||| |||||||| |||
Db      23905 TTGAGTGGGTCCTTGGCGGCTTCCAGCCCTCTGGTCCCAAGGGCCTGTAGCCACCATAAG 23846

Qy       1509 AAGAGAAGAATCCCTACAAAGAAGTGTACACGGACATGTGGGTGGAACCTGAGGCAGCTG 1568
              | ||||| ||||  ||||| ||||| ||||| |||||||||||||| |||||||||||||
Db      23845 AGGAGAACAATCTTTACAAGGAAGTTTACACCGACATGTGGGTGGAGCCTGAGGCAGCTG 23786

Qy       1569 CCTACGCACCACCTCCACCAGCCAAAAAGCCCCGGAAGAGCACAGCGGAGAAGCCCAAGG 1628
              | ||  | || || ||||||||||| || ||| | ||||||||        || | ||||
        23785 CTTATTCCCCGCCCCCACCAGCCAAGAAACCCAGAAAGAGCAC--------AACCTAAGG 23734

Qy       1629 TCAAGGAGATTATTGATGAGCGCACAAGAGAGCGGCTGGTGTACGAGGTGCGGCAGAAGT 1688
              |||||||||| |||||||| | | ||||| | || ||  |||| |||||||| || ||||
Db      23733 TCAAGGAGATCATTGATGAACACGCAAGATAACGACTCATGTATGAGGTGCGCCAAAAGT 23674

Qy       1689 GCCGGAACATTGAGGACATCTGCATCTCCTGTGGGAGCCTCAATGTTACCCTGGAACACC 1748
              |||  | ||| |||||||| || ||||| ||||||||||||||||| |||||||| ||||
Db      23673 GCCAAAGCATCGAGGACATTTGTATCTCATGTGGGAGCCTCAATGTCACCCTGGAGCACC 23614

Qy       1749 CCCTCTTCGTTGGAGGAATGTGCCAAAACTGCAAGAACTGCTTTCTGGAGTGTGCGTACC 1808
              | |||||| |||| ||||||||||| ||||| |||||||||||  |||||||||| ||||
Db      23613 CACTCTTCATTGGTGGAATGTGCCACAACTGTAAGAACTGCTTCTTGGAGTGTGCTTACC 23554

Qy       1809 AGTACGACGACGACGGCTACCAGTCCTACTGCACCATCTGCTGTGGGGGCCGTGAGGTGC 1868
              | ||||| ||||| |  |||||||||| ||| |||||||||||| |||| || || ||||
Db      23553 AATACGATGACGATGTGTACCAGTCCTTCTGTACCATCTGCTGTTGGGGGCGCGAAGTGC 23494

Qy       1869 TCATGTGCGGAAACAACAACTGCTGCAGGTGCTTTTGCGTGGAGTGTGTGGACCTCTTGG 1928
              ||||||| || ||||||||| |||||||||||||||| |||||||||||||| |||||||
Db      23493 TCATGTGTGGGAACAACAACCGCTGCAGGTGCTTTTGTGTGGAGTGTGTGGATCTCTTGG 23434

Qy       1929 TGGGGCCGGGGGCTGCCCAGGCAGCCATTAAGGAAGACCCCTGGAACTGCTACATGTGCG 1988
              ||||||| ||||||||||| ||||||||||||||||||||||| ||||||||||||||
Db      23433 TGGGGCCAGGGGCTGCCCAAGCAGCCATTAAGGAAGACCCCTGTAACTGCTACATGTG-- 23376

Qy       1989 GGCACAAGGGTACCTACGGGCTGCTGCGGCGGCGAGAGGACTGGCCCTCCCGGCTCCAGA 2048
                               | |||||||||| |  | ||||||||||| || || ||||| |
Db      23375 ----------------TGAGCTGCTGCGGAGAAGGGAGGACTGGCCTTCTCGACTCCACA 23332

Qy       2049 TGTTCTTCGCTAATAACCACGACCAGGAATTTGACCCTCCAAAGGTTTACCCACCTGTCC 2108
              ||||||| || ||||||||| ||||| |||||||||| || ||||||||||||||||| |
Db      23331 TGTTCTTTGCCAATAACCACCACCAGAAATTTGACCCCCCGAAGGTTTACCCACCTGTGC 23272

Qy       2109 CAGCTGAGAAGAGGAAGCCCATCCGGGTGCTGTCTCTCTTTGATGGAATCGCTACAGGGC 2168
              |||||||||||||||| |||||||  ||||| ||||| |||||||| || ||||||||||
Db      23271 CAGCTGAGAAGAGGAAACCCATCCACGTGCTATCTCTATTTGATGGGATTGCTACAGGGC 23212

Qy       2169 TCCTGGTGCTGAAGGACTTGGGCATTCAGGTGGACCGCTACATTGCCTCGGAGGTGTGTG 2228
              ||||||||||||||||| ||||||| || ||||||||||||||||| || |||||||| |
Db      23211 TCCTGGTGCTGAAGGACCTGGGCATCCAAGTGGACCGCTACATTGCTTCCGAGGTGTGCG 23152

Qy       2229 AGGACTCCATCACGGTGGGCATGGTGCGGCACCAGGGGAAGATCATGTACGTCGGGGACG 2288
              ||||||||||||| |||||||| |||||||||||||| |||||||||| |||||||||||
Db      23151 AGGACTCCATCACCGTGGGCATAGTGCGGCACCAGGGAAAGATCATGTTCGTCGGGGACG 23092

Qy       2289 TCCGCAGCGTCACACAGAAGCATATCCAGGAGTGGGGCCCATTCGATCTGGTGATTGGGG 2348
              || ||||| ||||||||||||||||||||||| ||||||||| | |||||||||||||||
Db      23091 TCTGCAGCATCACACAGAAGCATATCCAGGAGAGGGGCCCATACAATCTGGTGATTGGGG 23032
```

```
Qy       2349 GCAGTCCCTGCAATGACCTCTCCATCGTCAACCCTGCTCGCAAGGGCCTCTACGAGGGCA 2408
              ||||||| |||||||||||||||||| |||||||||| |||||||| || ||| ||||||
Db      23031 GCAGTCCTTGCAATGACCTCTCCATCATCAACCCTGCCCGCAAGGGACTTTACAAGGGCA 22972

Qy       2409 CTGGCCGGCTCTTCTTTGAGTTCTACCGCCTCCTGCATGATGCGCGGCCCAAGGAGGGAG 2468
              ||||| | ||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      22971 CTGGCAGCCTCTTCTTTGAGTTCTACCACCTCCTGCATGATGCGCGGCCCAAGGAGGGAG 22912

Qy       2469 ATGATCGCCCCTTCTTCTGGCTCTTTGAGAATGTGGTGGCCATGGGCGTTAGTGACAAGA 2528
              | ||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||
Db      22911 AGGATCGCCCCTTCTTCTGGCTCTTTGAGAAGGTGGTGGCCATGGCCGTTAGTGACAAGA 22852

Qy       2529 GGGACATCTCGCGATTTCTCGAGTCCAACCCTGTGATGATTGATGCCAAAGAAGTGTCAG 2588
              ||||||||||||||||||| ||||| ||| |||||||||||||  ||||||||||||| |
Db      22851 GGGACATCTCGCGATTTCTTGAGTCTAACGCTGTGATGATTGACTCCAAAGAAGTGTCTG 22792

Qy       2589 CTGCACACAGGGCCCGCTACTTCTGGGGTAACCTTCCCGGTATGAACAGGCCGTTGGCAT 2648
              |||||||||||| ||| |||||||||||||||||||| || ||||||||||| |||||||
Db      22791 CTGCACACAGGGTCCGTTACTTCTGGGGTAACCTTCCTGGCATGAACAGGCCATTGGCAT 22732

Qy       2649 CCACTGTGAATGATAAGCTGGAGCTGCAGGAGTGTCTGGAGCATGGCAGGATAGCCAAGT 2708
              ||||| |||||||||||||||||||||| ||||||||||| |||||||| | ||||| ||
Db      22731 CCACTATGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAACATGGCAGAACAGCCAGGT 22672

Qy       2709 TCAGCAAAGTGAGGACCATTACTACGAGGTCAAACTCCATAAAGCAGGGCAAAGACCAGC 2768
              |||||||||||||||||||||| || |||||||||||| |||||||||||||||||||||
Db      22671 TCAGCAAAGTGAGGACCATTACCACCAGGTCAAACTCCGTAAAGCAGGGCAAAGACCAGC 22612

Qy       2769 ATTTTCCTGTCTTCATGAATGAGAAAGAGGACATCTTATGGTGCACTGAAATGGAAAGGG 2828
              |||| |||||||||||||||||||| ||  ||||| | |||||||||||||||||||  |
Db      22611 ATTTCCCTGTCTTCATGAATGAGAAGGAAGACATCCTCTGGTGCACTGAAATGGAAACAG 22552

Qy       2829 TATTTGGTTTCCCAGTCCACTATACTGACGTC 2860
              | ||||| || ||    |    ||| || | |
Db      22551 TGTTTGGCTTACCCTGTCCTACTACAGAAGCC 22520
```

```
RESULT 11
AF176228
LOCUS       AF176228                4267 bp    mRNA    linear   PRI 26-OCT-1999
DEFINITION  Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA,
            complete cds.
ACCESSION   AF176228
VERSION     AF176228.1  GI:6118091
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4267)
  AUTHORS   Xu,G.-L., Bestor,T.H., Bourc'his,D., Hsieh,C.-L., Tommerup,N.,
            Bugge,M., Hulten,M., Qu,X., Russo,J.J. and Viegas-Pequignot,E.
  TITLE     Chromosome instability and immunodeficiency syndrome caused by
            mutations in a DNA methyltransferase gene
  JOURNAL   Nature (1999) In press
REFERENCE   2  (bases 1 to 4267)
  AUTHORS   Xu,G.L. and Bestor,T.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) Genetics and Development, Columbia
            University, 701 West 168 St., New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1..4267
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="20q11-q13"
     gene            1..4267
                     /gene="DNMT3B"
     CDS             265..2802
                     /gene="DNMT3B"
                     /note="mutated in human chromosome instability and
                     immundeficiency disease known as ICF syndrome; contains
                     alternative 5' exon 1B"
```

```
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3B"
/protein_id="AAF04015.1"
/db_xref="GI:6118092"
/translation="MEPSPEPSPLESMKGDTRHLNGEEDAGGREDSILVNGACSDSS
DSPILEAIRTPEIRGRRSSRLSKREVSSLLSYTQDLITGDGDGEDGGSDTPVMRKL
FRETRTRSESPAVRTRNNSVSSRERHRPSPRSTRGRQGRNHVDESPVEFPATRSLRR
RATASAGTPMPSPSSYLTIDLTDDTEDTHGTPQSSSSTPEYARLAQDSQQGMESPQVE
ADSGDGDSSEYQDGKEFGIGDLVWGKIKGFSWMPAMVVSWKATSKRQAMSGMRWVQNF
GDGKFSEVSADKLVALGLFSQHFNLATFNKLVSYRKAMTHALEKARVRAGKTFPSSPG
DSLEDQLKPMLEWAHGGFKPPGIEGLKPNNTQPENKTRRRTADDSATSDYCPAPKRLK
TNCYNNGKDRGGEDEDQSREQMASDVANNKSSLEDGCLSCCRKNPVSFHPLEFGGLCQYC
RDRFLELFYMYDDDGYQSYCTVCCEGRELLLCSNTSCCRCFCVECLEVLVGTGTAAEA
KLQEPWSCYMCLPQRCHGVLRRRKDWNWRLQAFFTSDYGLEYEAPKLYPAIPAARRRP
IRVSLFDGIATGYLVLKELGIKVGKYVASEVCEESIAVGTVKHEGNIKIYVNDVRNIT
KNIEEWGPDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYHLLNYSRPKEGGDDR
PFWMFENVVAMKVGDKRDISRFLECNPVMIDAIKVSAAHRARYFWGNLPGMNRPVIA
SKNDKLELQDCLEYNRIAKLKKVQTITTKSNSIKQGKNQLFPVVMNGKEDVLWCTELE
RIFGFPVHYTDVSNMGRGAROKLLGRSMSVPVIRHLFAPLKDYFACE"

Query Match      18.4%;   Score 791.8;  DB 9;   Length 4267;
Best Local Similarity   64.0%;   Pred. No. 1.1e-151;
Matches 1248;  Conservative   0;  Mismatches 667;  Indels  36;  Gaps  2;

BASE COUNT   1069 a   1098 c   1147 g   953 t
ORIGIN
```

(The remainder of the page consists of a nucleotide sequence alignment between Query (Qy) and Database (Db) sequences, with alignment match bars and base-position numbers ranging from 887 to 2816, arranged in paired Qy/Db rows.)

```
RESULT 12

LOCUS       AF068627            4135 bp    mRNA    linear   ROD 06-DEC-1999
DEFINITION  Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF068627
VERSION     AF068627.2  GI:6449471
KEYWORDS    .
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4135)
  AUTHORS   Okano,M., Xie,S. and Li,E.
  TITLE     Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
  JOURNAL   Nat. Genet. 19 (3), 219-220 (1998)
   MEDLINE   98324766
   PUBMED    9662389
REFERENCE   2  (bases 1 to 4135)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th street,
            Charlestown, MA 02129, USA
REFERENCE   3  (bases 1 to 4135)
  AUTHORS   Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK    Sequence update by submitter.
COMMENT    On Nov 18, 1999 this sequence version replaced gi:3327981.
FEATURES             Location/Qualifiers
     source          1..4135
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /map="84.0 cM"
     gene            1..4135
                     /gene="Dnmt3b"
     misc_feature    1..4135
                     /gene="Dnmt3b"
                     /note="similar to EST sequences deposited in GenBank
                     Accession Numbers AA116694, AA119979, AA177277, AA210568,
                     AA407106, and AA575617"
     CDS             269..2788
                     /gene="Dnmt3b"
                     /function="de novo DNA methylation"
                     /note="alternatively spliced product; contains Cys-rich
                     region; C-terminal region is similar to corresponding
                     region of DNA cytosine-5 methyltransferase 3A"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3B2"
                     /protein_id="AAC40179.2"
                     /db_xref="GI:6449472"
                     /translation="MKGDSRHLNEEDGASGYEECIIVNGNFSDQSSDTKDAPSPPVLE
                     AICTEPVCTPEPTPGRGRSSSRLSKREVSSLLNYTQDMTGDGDRDDEVDDGNGSDILMPK
                     LTRETKDIRTRSESPAVRTRHSNGTSSLERORASPRITRGRGRIHVQEPVEEPATR
                     SRRRASSSASTPWSSSASVDEMEEVTPKSVSTPSVDLSOQDGDQEGMDTTQVDAESRD
                     GDSTEYQDDKEFGIGDLVWGKIKGFSMWPAMVVSWKATSKRQAMPGMRWVQWFCDGKF
                     SEISADKLVALGFSQOHFNLATFNKLVSYRKAMYHTLEKAVRAGKTFSSSSPGESLED"
```

Db   2817 AAATGGAAAGGGTATTTGGTTTGGTTTCCCAGTCCACTATACTGACGTCTCCAACATAGAGCCGCCT 2876
              | |||| | ||||| ||| |||||||||||||||||||||||||||||||||||||
Qy   2651 AGCTCGAAAGGATCTTTGGCTTCCTTGCTGGACTCACACACAGCGTGTCCAACATGGGCGCCCGTG 2710

Qy   2877 TGGCGAGGCGAGGACTGCCTGGGCCGGTCATGGAGGCGTGCCAGTCATCCGCCACCTCTTCG 2936
              || || ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2711 GTGCCCCGCCAGAAGCTGCTGGGAAGGTCCTGGACGCGTGCCTGCCTGTGCATCCGACACCTCTTCG 2770

Qy   2937 CTCCGGCTGAAGGAGTGATTATTTGCGTGTGTGTA 2967
              | ||||||||||||||||||| || |||||||| ||
Db   2771 CCCCTCTGAAGGGACTACTTTGCCATGTGAATA 2801

```
BASE COUNT  1043 a  1048 c  1106 g  938 t
ORIGIN
```

Query Match
Best Local Similarity 18.4%;
Matches 1262: Conservative 64.6%; Pred. No. 5.9e-151; DB 10; Length 4135;
                           0; Mismatches 648; Indels 45; Gaps 4;
                                    Score 788.2;

```
QLKPMLEWANGGFKPTGIIEGLKPNKKQPENKSRRRTTNDSAASESPPPKRLKTNSYGG
KDRGEDEESRERMASEVYNNKGNLEDRCLSCRKKNPVSFHPLFEGGLCQSCRDRFLEL
FYWDEDGYQSYCYVCCEGRELLLCSNYSCCRCGFCVECLEVLVGAGTAEDAKLQEPWS
CYMCLPQRCHGVLRRRKDWNMRLQDFFTIDPDLEEFEPPKLYPAIPAAKRRRIRVLSL
FDGIATGIYLVLKELGIKVEKYIASEVCAESIAVGTVVKHEGQIKYVNDVRKITKRNIEE
WGPFDLVIGGSPCNDLSNVNPARRGLYEGTGRLFFEFYHLLNYTRPKEGDNRPFFWMF
ENVVAMKVNDKKDIJSRFLACNPVMIDAIKVSAAHRARYFWGNLPGNRPVWASKNDKL
ELQDCLEFSRTAKLIKKVQITTVKSNSIROGKNQLFPVVWNGKDDVLMCTELERIFGEP
AHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE
```

Qy   1016 GAGCCCGTGGGTGGCGTCCGATGCTGGGGACAAGAATGCCACCAAAGCAGGCCGATGAGCAGCCA 1075
              |||||||||||||| | |||||||||||||||||||||| ||||||||||||| |||||||||||||
Db   875 GATCAGGAGGGTGATTGGATCACCACACAGGTTGGATCAGAGAGAGCAGAGATGGAACAGCACA 934

Qy   1076 GAGTACGAAGGACCGGGGGCGTTTGGCATTGGGAGCTGGTTGGTGGGGGGAAACTCCGGGC 1135
              ||||| ||||| ||||| |||||||||||||||||| |||||||||||||| || ||||
Db   935 GAGTATCAGGATGATGAAAGAGTTTGAAATAGGTGACCTCGTGTGTGGGGAANGATCAAGGGC 994

Qy   1136 TTCTCCTGGTGGGCCAGGCCCGGCCATTGTGTCTTGGTTGGTGGGATGACGGGCCGGAGCCAGGCAGCT 1195
              ||||||||||| | |||||||||||||||||||||||| ||| || ||||||||||||| |||||||
Db   995 TTCTCCTGGTGGCCTGCCATGGTGGTGTCTTCTTGGAAAGCCACACTCCAAAGCGACACAGCCATG 1054

Qy   1196 GAAGGCACCCCGCTGGGTCATGTGGTTCGGAGACCGGCAAATTCTCAGTGGTGTGTGTTGTTGAG 1255
              | |||||||||||||||||| |||| ||||||| ||||||||||||||||| || |||||| |||
Db   1055 CCCGGAATGCGCCTGGGTACAGTGGTTTGGGTTGTGATGGCCAAGTTTCTGAGATCTGCTGAC 1114

Qy   1256 AAGCTGAATGGCCGGCTGAGCTCGTTTTGCAGTGCGTTCCACCAGGCCCACCTACAACAACAG 1315
              ||| ||||| | | |||| ||||| ||||||||||||| |||| ||| |||||||||||| ||| |
Db   1115 AAACTGGTGGCTCTGGGGCTGTTCAGCCAGCACTTTAATCTGGCTACCTTCAATAAGCTG 1174

Qy   1316 CCCATGTACCCGCAAAGCCATCTACGAGGTCCTGCAGGTGGCCCACAGCCGCGGGGGAAG 1375
              || | | |||||||||| ||||| ||||| ||| ||| ||||||||||||| || ||||||
Db   1175 GTTTCTTATAGGAAGGCCATGTACCACACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAG 1234

Qy   1376 CTGTTCCCGGTGTGCACGACAGCGGATGAGAGTCACNCGTGCCCAAGGCCGTGGAGGTGCCAG 1435
              ||||||||||||| |||| || | |||||||||| || |||| || |||| || |||| ||
Db   1235 ACCTTCTC---------------CAGCAGTCCTGGAGACTCACTGGAGGAACCAG 1273

Qy   1436 AACAAGCCCATGATTCAAATGGGCGCCTGGGGCCGGCTTCCACCCTTCTGGCCCCTAAGGGCCTG 1495
              | ||||||| || ||||||||||||||| ||||| ||||| |||||||||||||| |||||
Db   1274 CTGAAGCCCATGTCTGGAGTGGGCCCACGGTGGCTTCAAGCCTACTGGGAATCGAGGGCCTC 1333

Qy   1496 GAGCCACCAACAAAGAGAGAGAGAGAAATCCCTACAAAAGAGAGTGTACACGGACAATGTGGGTGGCC 1555
              ||||||||||| ||| | |||||||| |||||| ||||| || |||||| ||||| || |||||||
Db   1334 AAACC---------CAACAAGAAGCAACCAGAGAACAAAAAGTCGAAGACGAACAACCAATGAC 1387

Qy   1556 CCTGAGCGCAGCTGCCTACGCACCACCTCCACACCAGCCCAAAAAGCCCCGAAAGAGCGACACAGCG 1615
              ||||||||||||||||||| |||||||| |||| ||||||||||||| |||||||||| |||||
Db   1388 TCTGCGTCGCTTCTGAGTCCCCCCCCACCCAAGCGCCTCAGACAAATAGCTATGCGGGAAG 1447

Qy   1616 GAGAAGCCCAAGGGTCAAAGGAGAGATTATTGATGAGGGCCACAAGAGAGAGCGGCTGTGTACGAG 1675
              ||||||||||||||| |||||||||||||||||||| |||||||||||||||| || |||||||
Db   1493 GTCACCAAACAACAAGGGCAAATCTGGAAGACCGCGTGTTTGTCCGTGGAAGAAGAAACCCT 1552

Qy   1676 GTGCGGCAGAAGTGCCGGAACATTGAGGACATCTGCATCTCCTGTGGGAGGCCTCAATGTT 1735
              | |||| |||| |||| || ||||||||||||||||||| || || |||||||||||||||||
Db   1493 GTCACCAAACAACAAGGGCAAATCTGGAAGACCGCGTGTTTGTCCGTGGAAGAAGAAACCCT 1552

Qy   1736 ACCCTGGAAACACCCCCTCTTCGTTGGAGGAATGTGCCAAAAACTGCAAGAAACTGCTTTCTG 1795
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||||| ||
Db   1553 GTGTGTCCTTCCACCCCCTCTTTGAGGGTGGCGTCGTCAGAGTTGCCGGGATCGGCTTCCTA 1612

Qy   1796 GAGTGTCCGTACCAGTACGAGACGACGGGCTACCAGTCCACCATCTGCTGTGGG 1855
              ||| ||||| || ||||||| |||| ||||| ||||||||||||||||||||
Db   1613 GAGCTCTTCTCACATGTATGAGAGGACCGGCTATCAGTCCTACTGCACCACCGTGTCCTGTGAG 1672

Qy   1856 GGCCGTGAGGTGCTCATGTGCGGGAAACAACAAACTGCTGCAAGGTGCCTTTGCGTGGGAGTGT 1915

Db   1673 GGCCGTGAACTGCTGCTGTGTGCAGTAACACAAGCTGCTGCAGATGGCTTCTGTGTGGAGTGT 1732
Qy   1916 GTGGACCTCTTGGTGGGTGGGGGCCCGGGGGCTTGCCCAGGCAGCGCCATTAAGGAGAAGACCCCTGGAAC 1975
Db   1733 CTGGAGGTGCTGTGGTGGGCGCAGGCACAGTCGAAGGATGCCAAGCTGCAGGAACCTGGAGC 1792
Qy   1976 TGCTACATGTGCGGGGCACAAGGGTACCTTACGGCGTGCTGCTCGCGGACGAGGACTGGCCC 2035
Db   1793 TGCTATATGTGCCTCCCTCAGGCGCTGCCATGGGGTCCTCCGACCGCAGGAAGATTGGAAC 1852
Qy   2036 TCCCGGCTCCAGATGTTCTTCGCCTAATTAACCACGACC---AGGAATTTGACCCTCCAAAG 2092
Db   1853 ATGGCGCCTCCAAGACTTCTTCACTACTCGATCCTGAACTGGAAGAATTTGACCACCCAAG 1912
Qy   2093 GTTTACCCACCCTGTCCCAGCTGAAGAGGAAGCCCATCCGGGTGCTGTCTCTCTTTGAT 2152
Db   1913 TTGTACCCAGCAATTCCTGCACCCAAAAAGGAGGCCCATTAGAGTCCTGTCTCTGTTTGAT 1972
Qy   2153 GGAATCGCTACAGGGGGCTCCTGGTGCTGGAGCGACTTCAGGTGGACCGCTACATT 2212
Db   1973 GGAATTGCAACGGGGTACTTGGTGCTCAAGGAGTTGGGTATTAAAGTGGAAAAGTACATT 2032
Qy   2213 GCCTCGGAGGTGTGTGAGGACTCCATCACGGTGGGCACTGTCGCGGACCCAGGGGAGATC 2272
Db   2033 GCCTCCGAAGTCTGTGCCAGAGTCCATCCGTGTGGGAACTGTTAAGCATGAAGGCCAGATC 2092
Qy   2273 ATGTACGTCGGGGACGTCCGAGGTCACACGAGACCATATCCAGGAGTGGGGGCCCAATTC 2332
Db   2273 CGCCCCAAGGAGGGGCGCACACCGTCCAATTCTTCGTGGATGTTCGACAATGTTGTGGCCATG 2332
Qy   2393 GGCCTCTACGAGGGCACTGGGGCCCGGGCCTCTTCTTCTTTGAGTTCTACCTGCTCCATGATCG 2452
Db   2213 GGTTTATATGAGGGCACAGGGAAAGGCTCTTCTTCTTCTGAGTTTTACCACTTGCTGAATTATACC 2272
Qy   2513 GGCGGTTAGTGACAAGAGGACAATCTCGCGATTTCTCGAGTCCAACCCTGTGATGAATTGAT 2572
Db   2513 AAAGTGAATGACAAGAAGACAATCTCAAGAATTCCTGGCAATGTAACCCCAGTGATGATCGAT 2392
Qy   2573 GCCAAAGAAGGTGTCACGTGCCACACAGGGGCCCGGCGGTAAACCTTCCCGGTATG 2632
Db   2573 GCCATCAAGGTGTCTCGTGCTCCACTCACAGGGACGTGTTCCAACATGGGC 2632
Qy   2633 AACAGGCCGTTGGCATCCACTGTGAATGATGAAGCTGAGCTGCAGAGTGTCTGGAGCAT 2692
Db   2453 AACAGGCCCGTGATGGCTTCAAAGAAATGATTAAGCTCGAGCTGCCAGGACTCGGAGTTC 2512
Qy   2753 GGCAGGATAGCCAAGTTCAGCAAAGTGAGGACCACCATTACTACGAGGTCAAACTCCATAAAG 2812
Db   2513 AGTAGGACACGACAAAGTTAAAGAAAGTGCAGACAATAACCAAGTCGAACTCCATCGA 2572
Qy   2813 ACTGAAAATGAAAAAGGGTATTTGGTTTCCCAGTCCCACTATACTGACGTCTCCAACATGAGC 2872
Db   2633 ACTGAGCTCGAAAAGAGAATCTTCCGGCTTCCCTGCTCACTCACACGACGTGTCCAACATGGGC 2692
Qy   2873 CGCTTGGCGAGGCAGAGAACGTGCTTGGCGGGTCATGGAGCCGGTGCCAGTCATCATCCGCCACCTC 2932
Db   2693 CGCGGCGCCCGTCAGAGAGCTGCTCGGGAGGTCTCGGAGTCTACCGGTTCATCAGAGACACCTG 2752
Qy   2933 TTCGCCTCCGCTGAAGGAGAGTATTTTGCGTGTGTGTA 2967

Db   2753 TTTGCCCCCTTGAAGGACTACTTTGCCTGTGAATA 2787

Query Match   18.4%; Score 788.2; DB 10; Length 4163;
Best Local Similarity   64.6%; Pred. No. 5.9e-151;
Matches 1262; Conservative   0; Mismatches 648; Indels 45; Gaps 4;

Qy   1016 GAGCCCGTGGGGTCCCAATGCTCGGGGACAAGAAGTGCCACCAAAGCACGGCGATGACGAGCCA 1075
Db   903 GATCAGGAGGGTATGGATACCACACAGAGTGGATCCAGAGAGCACGCATAATGGAGACAGCACA 962
Qy   1076 GAGTACGAGGAGGCGCGGGGCGCCGGGGCTTTGGCATTGGGGAGCGCTGGTCTGTGGGGAAACTGCCGGGGC 1135

[The remainder of the page consists of a dense nucleotide sequence alignment between query (Qy) and database (Db) sequences, printed sideways, with position numbers ranging from approximately 963–2153 (query) and 1022–2815 (database) and alternating Qy/Db rows of DNA sequence and match bars.]

```
REFERENCE    2 (bases 1 to 4278)
AUTHORS      Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
             Shen,Y.
TITLE        Direct Submission
JOURNAL      Submitted (17-MAY-1999) Department of Biochemistry and Molecular
             Biology, Institution of Basic Medical Sciences, Chinese Academy of
             Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
             P.R.China
FEATURES             Location/Qualifiers
     source      1..4278
                  /organism="Mus musculus"
                  /strain="KM"
                  /db_xref="taxon:10090"
                  /dev_stage="8-9 day old embryo"
     gene        1..4278
                  /gene="Dnmt3b"
     CDS         412..2931
                  /gene="Dnmt3b"
                  /function="methylates cytosine in DNA"
                  /note="alternatively spliced"
                  /codon_start=1
                  /product="DNA cytosine-specific methyltransferase isoform
                  2"
                  /protein_id="AAF74516.1"
                  /db_xref="GI:8347120"
                  /translation="MKGDSRHLNEEGEASGYEECIIVNGNFSDQSSDTKDAPSPVLE
                  AICTEPVCTPETGRSRSSSRLSKREVSSLLNYTQDMTGDGRDDEVDDGNGSDILMPK
                  LTRETKDYRTRSESPAVRTRHNSNGTSSLERQRASPRITRGRQRHHVQEYPEPATR
                  SRRRASSSASTPWSSPASVDFMEEVTPKSVSTPSVDLSQDGDEGMDTTQVDAESIY
                  GDSTEYQDDKEPGIGDLVWGKIKGFSWWPANVVSWKATSKRQAPMGMRWVQWFGDGKF
                  SEISADKLVALGLFSQHFNLATFNKIVSYRKAMYHTLEKARVARAGKTSSSPGESLED
                  QLKPMLEWAHGGKFPTGIELKPNKQPENKSRRTNDSAASESPPRKLKTNSYGG
                  KDRGEDEESRERMASEVTNNKGNLEDRCLSCGKKNPVSFHPLFEGGLCQSCRDRFLEL
                  FYMYDEDGYQSYCYCVCCEGRELLLCSNTSCCRCFCVECLEVLVGAGTAEDAKLQEPWS
                  CYMCLPQRCHGVLRRRKDWNMRLQDFFTTDDLEEFEPPKLYPAIPAAKRRPIRVLSL
                  FDGIATGYLVLKELGIKVEKYIASEVCAESIAVGTVKHEGQIKYVNDVRKITKNIEE
                  WGPFDLVIGGSPCNDLSNVNPARKLIYEGTGRLFFEFYRLLNYTRPKEGDNRPFFWMF
                  ENVVAMKVNDKKIDISRFLACNPVMIDAIKVSAAHRARYFWGNLPGMNRPVMASKNDKL
                  ELQDCLEFSRTAKLKKVQTITKSNSIRQGKNQLFPVVMNGKDVLWCTELERIFGFP
                  AHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"
BASE COUNT   1055 a   1113 c   1154 g    956 t
ORIGIN
```

```
Query Match                18.3%;        Score 786.6; DB 10;   Length 4278;
Best Local Similarity  64.5%;        Pred. No. 1.3e-150;
tches 1261; Conservative   0; Mismatches  649;  Indels   45;   Gaps   4;
```

```
Qy    1016 GAGCCCGTGGGGTCCGATGCGTGGGGACAAGAGAATGCCACCAAAGCAGGCGATGACGAGCCA 1075
           | ||  |  ||||         |  ||||||||   |||| ||||||||||||||||||| ||||||||
Db    1016 GATCAGGAGGGTATGGGATACCACACAGGTGGATGCAGAGAGACATATATGAGACACACA 1077

Qy    1076 GAGTACGAGGACGACGGCCGGGGGCTTTGGCATTGGGGAGCTGGTGGGGAAACTGCGGGGC 1135
           |||||    |       | |  || |||||| |||||| |||    |  | |||||
Db    1078 GAGTATCAGGATGATAAAAGAGTTTGGAATAGGTGACCTCGTGTCGGGGAAGATCAAGGGC 1137

Qy    1136 TTCTCCTGGTGGCCAGGCCCATTGTTGTCTTCGGTTGGATGACGGGCGGGACCCGACAGCT 1195
           || |  ||||||||||    |||||||| |||| |  |||  |||||||||| |||||
Db    1138 TTCTCCTGGTGGCCCTGCCATGGTGGTCTCCTGGAAAGCCACCTCCAAGCGACAGGCCATG 1197

Qy    1196 GAAGGCACCGCCGCTGGGGTCATTGGTTCCGGAGACGGCAAATTCTCAGTGGTGGTGGTGGAG 1255
           |  ||  ||||||||  |||||||||  || |||| ||||||||||||||||||| ||||||  |  ||
Db    1198 CCCGGAATGCCGCTGGGGTACAGTGGTTTGGTGATGGCAAGTTTTCTGAGATCTCTGCTGAC 1257

Qy    1256 AACCTGATCCCGCTGAGCTGGTTTTCCAGTCGCGTTCACCAGGCCCACGTACAACAAGCAG 1315
           ||   | |||||  || ||||   | | ||||||| || |  |||||||| || ||||||
Db    1258 AAACTGGTCGGGCTCTGGGGCGTGTTCACCCAGCACTTTAATCTGGCTACCTTCAATAAGATCG 1317

Qy    1316 CCCATGTACCGCAAAGCCATCTACGAGGTGGCCAGGGTAGCGACGACGCCGGGGGAAG 1375
           ||    ||| || |  |  ||||  |  |||||||||  || || |||| ||||  | |
Db    1318 GTTTCTTATAGGAAGGCCACCAGGAATACCACACTCTGGAGAAAGCCAGGGTTCGAGCTGGCCAAG 1377

Qy    1376 CTGTTCCCGGTGTGCCCACGACGCGATGAGCGTGACACTGCCAAGGCCGTGGAGTTGCAG 1435
           || | ||
Db    1376

Db    1378 ACCTTCTC------------------------CAGCAGTCCTTGGAGAGTCACTGGGGAGGACCAG 1416

Qy    1436 AACAAGGCCCATGATTGAAATGGGCGGGCCCTTTCCAGCCTTCTCGGGCGGCTTCTCGGGGACGAG 1495
           |||| | || || |  |||| | ||||||||||    |||||||  ||||||||  ||||||
Db    1417 CTGAAAGCCCATGCTGAGGTGGGCCCCACGGTGGCTTCAAGCCTACTGGGACGCAGGGGCTC 1476

Qy    1496 GAGGCACCAGGAAGAAGAAGAGAAGAATCCCTACAAAAGAAAGTGTACACGGACACATGTGGGTGGAA 1555
           ||||||  | | | || || |  ||||||||||||| || ||  |||  | || |||||||  ||
Db    1477 AAACC-------CAACAAGAAGCAACCAGAGAAACCAGAGAAACAAAAGTCGAAGAGCGCACAACCAATGAC 1530

Qy    1556 CCTGAGGCAGCTGCCTACGCACCACCTCCACCAGCCACCAAAAAAGCCCCGGAAGAGCACAGCG 1615
           || ||||||  || || || |||||| |  |||  |  || |||||||||||||
Db    1531 TCTGCTGCTTCTGCTGAGTGCCCCACCCCCCACCCAAGCGCCCTCAGACAAATAGCTCATGGCCGGGAAG 1590

Qy    1616 GAGAAGCCCAAGGTCAAGGAGAGATTATTGATGACGCGCACAGAAGAGAGCGCTGGTGTGTACGAG 1675
           ||||||       ||||  |||||||  |||||
Db    1591 GACCCGAGGGGGAGGGGAGGATGAGGAG------------AGCCGAGAACGGAGGGATGGCTTCTGAA 1635

Qy    1676 GTGCGGGCAGAAGTGCCGGAACATTGAGGACATCGCAATCTGGACATTCCCGTGGGGGAGGCCTCAATGTT 1735
           ||                 |||||||||       || |||  |||||||||
Db    1636 GTCACCAACAACAACAAGGGCAATCTGGAAGAGCACCCGTGTTTGCCTTGTCCTGTGGAAGAGAACCCT 1695

Qy    1736 ACCCTGGACAACACCCCCTCTTCGTTGGGAAGATGTGCCAAAACTGCAAGAACTGCTTTTCTG 1795
           |||||||||||| ||||||||||||  ||  |||||||||||  |||||||||||||
Db    1696 GTGTGCCTTCCACACCCCTCTTTGAGGGTGGCGTCTGTCAGAGTTGCCGGGATCGCTTCCTA 1755

Qy    1796 GAGTGTGCGTACCAGTACGACGACGACGGCCTACCAGTCCTTGGGGGCTCTGTCACCATCTGCCGTGGG 1855
           |||| |||  | |||||  || ||  ||||||  ||||  ||||  || |  ||||
Db    1756 GAGCTCTTCTACATGTATGATGAGGACGGGTATCAGTCCTACTGCACCGTGTGTTGCGAG 1815

Qy    1856 GGCCGTGAGGTGCTCATGTGCGGAAACAACACATGCTGCAGGTGCTTCTTTGTCGTGGAGTGT 1915
           ||||| ||||| |  |  |||||  |       |||||||   ||||||
Db    1816 GGCCGTGAACTGCTCGCTGTGCGAGTAACACAAGCTGCTGAGAGTGCTTCTGTGTGGAGTGT 1875

Qy    1916 GTGGACCTCCTTGGTGCGGGACGAAGGGTCACCTACGGCTGCTGCCGGGGCTGCCCAGGCAG 1975
           |||||   ||||  |||  |  || |||  ||   | ||||||||   ||||||||
Db    1876 CTGGAGGTGCTGGTGGGCGCGGCGAGCACCGCTGAGGATGCCAAGCTGCAGGAGCCCTGGAGC 1935

Qy    1976 TGCTACATGTGCGGGCCACAAGGGTACCTACGGCGCCTGCCTGAGCTCCTGCCGGGCGCGCGAGC 2035
           | ||| ||||| |  |  ||||   |  ||| |||| |||
Db    1936 TGCTATATGTGCCTCCCCCAGCGCTGCCATGGGGTCCTCCGACCGGAGAAGGAAAGATTGGAAC 1995

Qy    2036 TCCCGGCTCCAGATGTTCTTCGGCTAATAACCACGACC------AGGAATTTGACCCTCCAAAG 2092
           ||  |||||||  || || |||   || |||||||||      ||||||||||||||||||
Db    1996 ATGCGCCTGCAAGACTTCTTCACTACTGATCTGGAAGAATTTGAGCCACCAAAG 2055

Qy    2093 GTTTTACCCACCTGTCCCACGCGTCGAGAGAGAGGAAGACCCGGCGTGTCTCGTGTCTTTTGAT 2152
           | |||||    |   ||  | |||  |||||||||  ||||||||  |||| |||| |||||
Db    2056 TTGTACCCACCAATTCCTGCAGCCAAAAGGAGGCCCAATTAGAGTGCTGTCTGCTGACTCGGTTTGAT 2115

Qy    2153 GGAATCGCTACAGGGCGTCCCTGGTGCTGAAGGAGTGGGGCATTCAGGTGGGCCGCTACATT 2212
           || ||  |  | |||| ||| || ||||| |  ||||| | |||||  |||||
Db    2116 GGAATTGCAAACGGGGTACTTGGTGCTGGTTAAAGTGGAAAAGTACATT 2175

Qy    2213 GCCTCGGAGTGTGTGAGGACTCCATCACGGTTGGCGGGCACCAGGGGAGATC 2272
           |||  ||||||||| ||  | ||||  | | |  |||  | |||||||||||
Db    2176 GCCTCCGAAGTCTGTGCAGAGTCCATCGCTGTGGGAACTGTTAAGCATGAAGGCCAGATC 2235

Qy    2273 ATGTACGTCGGGGACGTCCGCGCGTCACACAGACATATCCAGGAGTGGGGCCCCATTC 2332
           | | |    |||  |||||  | ||| |||  ||  ||||| |  |||||
Db    2236 AAATATGTCAATGACGTCCGGAAAAATCACCAGAAAATATTGAAGAGTGGGGCCCGTTC 2295

Qy    2333 AACCTGATCGTCGGGGGCAGTCCCTGCAATGACCTCTCCATCGTCAACCCTGCTCCCAAG 2392
           ||  || ||||  ||||| | || |||||||| | | ||| |  ||||||
Db    2296 GACTTGGTGATTGGTGGAAGCCCATGCAATGATCTCTCTAACGTTAACCCTGCCCGCAAA 2355

Qy    2393 GGCCTCTACGAGGGCACTGGCCGGGGCGGGGCCCTCTTCTTCTTTGAGTTCTACCGCCTGCATGATGCG 2452
           |||  ||| | |||  | |||    ||||
Db    2356 GGTTTATATGAGGGCACAGGAAGGCTCTTCTTCTTCGAGTTTTACCGACTTGCTGAATATACC 2415

Qy    2453 CTGTTCCCGGTGTGCCCACGACGCGATGAGCGTGACACTGCCAAGGCCGTGGAGTTGCAG 2512
           |||  ||||| | | |||| |||| ||| | ||| ||||  | ||
Db    2416 CGCCCCAAGGAGGGGGAGAATGATCGCCCCTTCTTCTGGATGTTTGGCCATG 2475
```

RESULT 15.

```
LOCUS       AF156488     4145 bp    mRNA    linear   PRI 05-SEP-1999
DEFINITION  Homo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)
            mRNA, complete cds.
ACCESSION   AF156488
VERSION     AF156488.1  GI:5823167
KEYWORDS    .
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4145)
  AUTHORS   Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.
            and Li,E.
  TITLE     Cloning, expression and chromosome locations of the human DNMT3
            gene family
  JOURNAL   Gene 236 (1), 87-95 (1999)
   PUBMED   10433969
REFERENCE   2  (bases 1 to 4145)
  AUTHORS   Xie,S. and Li,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
FEATURES             Location/Qualifiers
     source          1..4145
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between D20S182 and D20S106; 39.9-50.2 cM"
     gene            1..4145
                     /gene="DNMT3B"
     CDS             115..2676
                     /gene="DNMT3B"
                     /note="de novo DNA methyltransferase"
                     /codon_start=1
                     /product="DNA cytosine-5 methyltransferase 3 beta 1"
                     /protein_id="AAD53063.1"
                     /db_xref="GI:5823168"
                     /translation="MKDTRHILNGEDADGGREDSILVNGACSDQSSDSPPILEAIRTP
                     EIRGRSSSRLSKREVSSLLSYTQDLIGDGDEDGDGSDTPVMPKLFRETRTRSESPA
                     VRTRNNSVSSRERHRPSPRSTRGRGRGRNHVDESPVEFPATRSLRRRATASAGTPWPS
                     PSSYLTDLTDDTEDTHGPTPGGSSSTPYARLAQDSQQGWMESSPQVEADSGDGSSEYQ
                     DKREFGIGDLVWGKIKGFSWWPAMVVSWKATSKRQAMSGMHRWVQWFGDGKFSEVSADK
                     LVALGLFSQHFNLATFNRLVSYRKAMYHALEKARVARAGKTPPSSPGDSLEDQLKPMLE
                     WAHGGFKPTGIEGLKPNNTQPVVNKSKVRRAGSRKLESRKYENKTRRRTADDSATSDY
                     CPAPKRLKTNCYNNGKDRGDEDQSREQMASDVANNKSSLEDGCLSCGRKNPVSHPLF
                     EGGLCQTCRDFLEELFYMYDDDGYQSYCTVCCEGRELLLCGNTSCCRCFVECLELV
                     GTVAAEAKLQEPWSCYMCLPQRCHGVLRRRKDWNVRLQAFFTSDTGLEYEAPKLYPA
                     IPAARRRPIRVLSLFDGIATGYLVLKELGIKVGKYVASEVCEESIAVGTVKHEGNIKY
                     VNDVRNITTKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEYHLLNYS
                     RPKEGDDRPFFWMFENVVAMKVGDKRDISRFLECNPVMIDAIKVSAAHRARYFWGNLP
                     GMNRPVIASKNDKLELQDCLEYNRIAKLKKVQTIITTKSNSIKQGKNQLFPVVMNGKED
                     VLWCTELERIFGFPVHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"
BASE COUNT     1041 a   1083 c   1096 g    925 t
ORIGIN
```

```
Query Match      17.8%;  Score 764.4;  DB 9;  Length 4145;
Best Local Similarity   63.0%;  Pred. No. 4.5e-146;
Matches 1244;  Conservative  0;  Mismatches 706;  Indels 26;  Gaps 3;
```

```
Qy  2513 GGCGTTAGTGACAAGCAGGGACACTCGCGGATTTCTCGAGTCAAACCCTGTCGATGATTGAT 2572
Db  2476 AAAGTGAATGACAACGAAAAGACATCTCAAGATTCCGTGGCATGTAACCCAGTGATGATCGAT 2535
Qy  2573 GCCAAAGAAGTGTCAGCTGCACACAGGGCCCCTACTTCTGGCCGGTTATG 2632
Db  2536 GCCATCAAGGTGTCTGCTGCTCACAGGGCCGGTACTTCTGGGGTAACCTACCCGGAATG 2595
Qy  2633 AACAGGCCGTTGGCATCCACTGTGAATGATGAAGCTGCAGGAGTGTCTGGAGCCAT 2692
Db  2596 AACAGCCCGTGATCGGCTTCAAAGAATGATAAGCTCGAGCCTGCAGGACTGCCTGGAGTTC 2655
Qy  2693 GGCAGGATAGCCGAAGTTCAGCAAAGTGAGGACCACCATTACTACGAGGTCAAAACTCCATAAAG 2752
Db  2656 AGTAGGACACGCAAAGTTAAAGAAAGTGCAGACAATAACCACCAAGTCGAAACTCATCAGA 2715
Qy  2753 CAGGGCAAAGACCAGCGATTTTTCCTCTCTTCATGAATGAAGAAAGAGGACAATCTTATGGTGC 2812
Db  2716 CAGGGCAAAAAACCAGCTTTTCCCTGTAGTCATGAATGGCAAGGCACGGCAAATTTCTCAGTGGTTC 2775
Qy  2813 ACTGAAAATGGAAAAGGGTAATTGGTTTCCCAGTCCACTATACTGACGGTCTTCCAACACATCAGAGC 2872
Db  2776 ACTGAGCTCGAAAAGAATCTTCGGGCTTCCCTGCTCACTACACGGACGTGTCCAACATGGGC 2835
Qy  2873 CGCTTGGCGAGGGCAAGAGACTGCTGGGCCCGTCATGGAGGGTGCCCCGCACTTGCTGTGGGGGAAACTCGGGGCT 2932
Db  2836 CGCGGCGCCCCGGTCAGAAAGCTGCTGGGCCAGGTCGTTCCATTGGGGAGGTGTTGGGGGAAAACTGGCAGTTCAG 2895
Qy  2933 TTCGCTCCGCTGCAAGGGAGTGATATTTTTGGGTGGTGTGAATA 2967
Db  2896 TTTGCCCCCTTTGAAGGACTACTTTGCCTGTGAATA 2930
```

```
Qy  1017 AGCCCCGTGGGGTCCGATGCTGGGGACAAGAGAAATGCCACCAAAAAGCAGGCGATGACGAGCCAG 1076
Db   701 AGCAGGGGGGCCATGGAGTCCCCGCAGGTGGAGACACTGAGGACAGTTCAG 760
Qy  1077 AGTACGAGGACGAGCACCGCGGGGGCTTGTGGGGCATTGGGGACGTTGGTGTGGGGAAAAATCGGGGCT 1136
Db   761 AGTATCAGGATGGGAAGGAGTGTTGGAATAGGGGACCTTCGTGTGTGGGAAAGATCAAAGGGCT 820
Qy  1137 TCTCCTGGTGGCCAGCGCGCACATTGTGTCTTGGGTGGATGACGGGCCGGAGGCCGGAGCAGCTG 1196
Db   821 TCTCCTGGTTGGCCGCCATGTCGTGGTTTGGAAGGCCACCTCCAAGCGACAGCAGGCTATT 880
Qy  1197 AAGGCACCGCCTGGGGTCATGGTTCGGAGACGGGCAAAATTCTCAGTGGTTGGTTGTTGAGA 1256
Db   881 CTGGCATGCGGTGGGGTCCAGTGGTTGGGCGATGGCAAAGTTCTCCGAGGTCTCTCCGACA 940
Qy  1257 AGCTGATGCCCCTGAGCTCGTTTCTTGCAGTGCGTTCGTTCCACCAGGCCCACGTACACAACAAGCAC 1316
Db   941 AACTGGTGGCACTGGGGCTGTTCAGCCAGCACTTTAATTTGGCACCTTCAAATAAGCTCG 1000
Qy  1317 CCAATGTACCGCAAAGCCATCTACGAGGTCCTGCCAGGTGGGCCAGACGCCGCGCGGGGAAGC 1376
Db  1001 TCTCCTATCGAAAAGCCATGTACCATGCTCTGGAGAAAGCCAGAGTGGCCAGACTGGCCAAGA 1060
Qy  1377 TGTTCCCGGTGTGTGCCACGACAGCCGATGAGAGTGACACTGCCAAGGCC---GTGGAGGTGC 1433
Db  1061 CCTTCCCCAGCAGCCCTGAGACTCATTGGAGGAGCCAGCTGAAGCCCATGTTGGAGTGGG 1120
Qy  1434 AGAACAAGCCCATGATTGAAATGGGCCCTGGGGGCGTTCCAGCC---------- 1476
Db  1121 CCCACGGGGGCTTCAAGCCCACTGGGGATCGAGGGCTCAAACCCAACAACACGCAACCAG 1180
Qy  1477 -----TTCTGGCCCTAAGGGCCTGGAGCCACCAGAAGAGAGGAGAATCCCTACAAAGAA 1531
Db  1181 TGGTTAATAAGTCGAAGGTGCGTCGTGCAGGCAGTAGGAAATTAGAAT-CAAGGAAATAC 1239
Qy  1532 GTGTACACGGACGACATGTGGGTGGAAACCGTGAGGCCAGCTGCCTTACGACCACACTCCACCAGCC 1591
Db  1240 GAGAACAAGACTCGAACGCACAGGTGACCAGCCACACTTGACTACTGCCCCGCA 1299
Qy  1592 AAAAAAGCCCGCGAAGAAGCCACACGCGGAGAAGCCCAAGGTCAAGGAGATTATTGATGAGCGC 1651
Db  1300 CCCAAGCGCCTCAAAGACAGAAATTGCTATAAACAACGGACACCGAGGGGAATGAAGATCAG 1359
Qy  1652 ACAAGAGAGACGGCGGTGGTGTACGGAGGTCGGCGCAGAAGTGCCGGAACATTGAGGACACTGC 1711
Db  1360 AGCCGAGAACAAATGGCTTCAGATGTTGCCAACACACAAAGACCAGCCTGGAAAGATGGCTGT 1419
Qy  1712 ATCTCCCTGTGGGAGCCTCAATGTTACCCGTGGAACACCCCCCTCTTCGTTGGAGGAATGTGC 1771
Db  1360 AGCCGAGAACAAATGGCTTCAGATGTTGCCAACACACAAAGACCAGCCTGGAAAGATGGCTGT 1419
```

```
Db  1420  TTGTCTTGTGTGGCAGGAAAAACCCCGTGTCCTTCCACCCCTTCTCTTTGAGGGGGGGGGCTCTGT  1479

Qy  1772  CAAAACTGCAAGAAACTGCTTTTCTGGGACTGTGGCTGGCTACCAGTACGACGACGGCGGCTACCAG  1831
          | |                 | ||| ||||||| |||| ||     |  |||||  ||||||| |||

Db  1480  CAGACATGCCGGGATCGCGTTCCTTGACCTGTTTTTACATGTATGATGACGATTGGCTTATTCAG  1539

Qy  1832  TCCTACTGCACCATCTGTCTGTGGGGGGCCGTGAGGTGCTCATTGTGGGGAAACAAACTGC  1891
          |   |||||||     | |  |||| |  |  |||   ||  |  |||| |||| ||||

Db  1540  TCTTTACTGCACTGTGTTGTTGCTGGAGGGGCCGAGAGACCTGCTGGTGCAACACGAGCTGC  1599

Qy  1892  TGCAGGTGCTTTTTGGCGTGGGACTGTGTGTGGACCTCTCTTGGTGTGGGGGCCGGGGGCTGCCCAGGCA  1951
          |||||| ||||||||||| ||| | ||||| |||||  |||| ||||   |||  ||||| | ||| |||

Db  1600  TGCCGGTGTGTTTTTCGTGTTGTGTGGGAGTTGCCTGGAGTGGCACAGGCACACGGCCGAG  1659

Qy  1952  GCCATTAAGGAAGAACCCCTGGAACTGCTACATGTGCGGGGCACAAGGGTACCTTACGGGGCTG  2011
          |||                | |||  | |||||    |||| ||  |||| | |||||||| |||

Db  1660  GCCAAGCTTCAGGAGCCCTGGAGCTGCTACACAGTGCTCCCCGCACGGCCTGTTCATGGGCCTC  1719

Qy  2012  CTGCGGCGGCGGAGAGGACTGGCCCTTCCCCGGGCTACACAGGACTGTTCTTCGCTAATAAACCACGAC  2071
          | | ||  ||| ||||  |  |||| ||  |||  |||| ||   ||| |||||||   |  || || ||

Db  1720  CTCCGGGCGCGGAAGGACTGAAACGTGCGCCTCCCCGGCTCCAGATTGTCTCCCGCAGCCGGTGACACCGAC  1779

Qy  2072  CAGGAATTTGACCCTCCAAAGGTTTACCCACCTGTCCCAGCTGAGAGAGAGGAAGCCCATC  2131
          |  ||  |||||||   |||| ||| |||  ||| |||| |  ||| ||||||||| ||||

Db  1780  CTTGAATACGAAGCCCCCAAGCTGTACCCTGCCATTCCCGGAGTGTGTGAGGAGTCCATTGCTGTTGGAACC  1839

Qy  2132  CGGGTGCTGTCTCTCTTTGATGGAAATCGGCTACAGGGCCTCCTGGTGCTCGAAGGACTTGGGC  2191
          |  |||| || ||   ||| || |||  |||| |||  ||| ||| ||||||||| |||||||

Db  1840  CGAGTCCTGTCATTGTTTGATGGCATCCGACCAGAGGCTACCTAGTTCCTCAAAGAGTTGGGC  1899

Qy  2192  ATTCAGGTGGACCGCTACATTGCCTCCGGAGGTGTGTTGTGAGGACTCCATCACCGTGGGCATG  2251
          ||  ||||| |||||||| ||| |||  |||| |||||||  |||| |||| ||||||||||  |

Db  1900  ATAAAGGTAGGAAAAGTACGTCGCTTCTGAAGTGTGTGAGGAGTCCATTGCTGTTGGAACC  1959

Qy  2252  GTGCGGCACCAGGGGAAGATCATGTACGTCGGGGAGACGTCCGACGCGTCACACAGGAAGCAT  2311
          | | ||| ||||||||| |||  | || ||||| ||||  |||  || |  ||||||   ||

Db  1960  GTGAAGCACGAGGGGAAATATCAAATACGTGAACGACGTGAGGAACATCACAAAGAAAAAT  2019

Qy  2312  ATCCAGGAGTGGGGGCCCCATTCGATCTGGTGATTGGGGGCAGTCCCTGCAATGACCTCTCC  2371
          ||||  |||||| |||||||||  | || |||| |||  | || |||||||||||||||||

Db  2020  ATTGAAGAATGGGGCCCCATTTGACTTGGGTGATTGGCGGAAGCCCATCAACGATCTCTCA  2079

Qy  2372  ATCGTCAACCCTGCTCGCAGGGCCTCTACGAGGCCACTCGGCCGCGCTCTTCTTCTTTGAGTTC  2431
          |   ||||| ||| |  ||| |||   || | |||| ||  ||  | || ||||| |||| |||

Db  2080  AATGTGAATCCAGCCAGGAAAGGCCTGTATGAGGGTACAAGGCCGGCTCTTCTTCGAATTT  2139

Qy  2432  TACCGCCTCCTGCATGATGCCGCGGCCCAAGAGGAGGAGGATGAATCGCCCCTTCTTCTGCTC  2491
          |||||  || || ||| |||  |||   ||||||||| |||  |||| |||| ||||||||||

Db  2140  TACCACCTGCTGAATTACTCAGCCCCAAGAGGAGGGTGATGACCGGGATG  2199

Qy  2492  TTTGAGAATGTGGGTGGCCATGGGAGTTAGTGACAAGAGGGACAATCTCGCGATTTCTCGAG  2551
          || |||||||| |||||| |||| ||| ||  || ||||| ||||| ||||  |||| ||||

Db  2200  TTTGAGAATGTTGTTAGCCATGAAAGGTTGGCGACAGAGAGGACACATCTCACGGTTCCTGGAG  2259

Qy  2552  TCCAACCCTGTGATGATTGATCCCAAAGAAGTGTGCCACACAGGGCCCGCTACTTTC  2611
          |   |  | ||| ||||||  |||||||| || | ||||||||||||||  || |||

Db  2260  TGTAATCCAGTGATGAATTGATCCCATCAAAGTTTCTGCCTCACAGGGCCCGATACTTC  2319

Qy  2612  TGGGGTAACCTTCCCGGGTTATGAGAACAGGGCCGTTGGGCATCCACACTGTGAATAGTAGCTGGAG  2671
          | |||| |  |||| ||| | ||||| ||||| | ||| |||||| |  | ||  | |||||

Db  2320  TGGGGCAACCTACCCGGGGATGAACAGCGCCCGTGATAGCATCAAAGAATGATAAAACTCGAG  2379

Qy  2672  CTGCAGGAGTGTCTGCTGGAGCATGGCAAGGTTCAGCAAAGTGAGAGACCCATTACT  2731
          | ||||||| ||| |||||| ||| | ||||| | ||  || ||  || ||||| |||

Db  2380  CTGCAGGACTGCTTGGAATACAAATAGGATAGCCAAGTTAAAGAAAGTACAGACAAATAACC  2439

Qy  2732  ACGAGGTCAAAACTCCATAAAAGCAGGGCAAAGACCAGCACATTTCCTGTCTTCCATGAATGAG  2791
          | ||||| || ||  | | |||| |||  ||||||| ||| |||| || | | |||||| ||||

Db  2440  ACCAAGTCGAACTCGATCAAACAGGGGAAAAAACCAACTTTCCCCTGTTGTCAATGAATGC  2499

Qy  2792  AAAGAGGACATCTTATGGTGCACTGAAATGGAAAGGGTATTTGGTTCCCAGTCCCACTAT  2851
          ||||| | || | |  |||  ||||| || |||| || | |||| |||| |||||||

Db  2500  AAAGAGAGATGTTTGTGGTGCACTGAGCTCGAAAGGATCTTTGGCTTTCCTGTGCACTAC  2559
```

```
Qy  2852  ACTGACGTCTCCAACATGAGCGCCGCTTGCGAGGCGAGAGACTGCTGGGCCGGTCATGGAGC  2911
          || |||||| |||||||||| |||||||| || | |||||   ||||||||  ||||||||
Db  2560  ACAGACGTGTCCAACATGGGCGCCGCTGGGTGCCCAGCAGGCGGAAGCTTCCTGGAGC  2619

Qy  2912  GTGCCAGTCATCCGCCCACCTCTTCGCTCCGCTGAAGGAGGTATTTTGCGTGCGTGTA  2967
          |||||| || ||||| ||||||||||| || |||||||| || ||||| || ||
Db  2620  GTGCCTGTCATCCGACACCTCTTCGCCCCCTCTGAAGGACTACTTTTGCATGTGAATA  2675
```

THIS PAGE BLANK (USPTO)

RESULT 2

LOCUS   AF067972   3005 bp   mRNA   linear   PRI 12-FEB-2001
DEFINITION Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,
            complete cds.
ACCESSION   AF067972
VERSION   AF067972.2 GI:12746531
KEYWORDS   .
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3005)
  AUTHORS   Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.
            and Li,E.
  TITLE   Cloning, expression and chromosome locations of the human DNMT3
            gene family
  JOURNAL   Gene 236 (1), 87-95 (1999)
  MEDLINE   9365304
   PUBMED   10433969
REFERENCE   2 (bases 1 to 3005)
  AUTHORS   Xie,S., Okano,M. and Li,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th
            Street, Charlestown, MA 02129, USA
REFERENCE   3 (bases 1 to 3005)
  AUTHORS   Hata,K., Shirohzu,H., Sasaki,H. and En,L.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th
            Street, Charlestown, MA 02129, USA
REMARK   Sequence update by submitter
COMMENT   On Feb 12, 2001 this sequence version replaced gi:4927369.
FEATURES   Location/Qualifiers
     source   1..3005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p23; between D2S171 and D2S174; 48-50 cM"
     gene   1..3005
                     /gene="DNMT3A"
     CDS   230..2968
                     /gene="DNMT3A"
                     /function="de novo DNA cytosine methyltransferase"
                     /codon_start=1
                     /product="DNA cytosine methyltransferase 3 alpha"
                     /protein_id="AAD33084.2"
                     /db_xref="GI:12746532"
                     /translation="MPAMPSSGPGDTSSSAAEREEDRKDGEEQEEPRQKEERQEPSTT
                     ARKVGRPGRKRKHPVESGDTPKDPAVISKSPSMAQDSGASELLPNGDLEKRSEPQPE
                     EGSPAGGGKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPAEAGKEQKETNIES
                     MKMEGSRGRLRGGLGWESSLRQRPMPRLTFQAGDPYYISKRKRDEWLARWKREAEKKA
                     KVIAGMNAVEENQGPGESQKVEEASPPAVQQPTDPASPTVATPEPVGSDAGDKNATK
                     AGDDEPEYEDGRGFGIGELVWGKLRGFSWWPGRIVSWWMTGRSRAAEGTRWVMWFGDG
                     KFSVVCVEKLMPLSSFCSAFHQATYNKQPMYRKAIYEVLQVASSRAGKLFPVCHDSDE
                     SDTAKAVEVQNKPMIEWALGGFOPSGPKGLPPEEEKNPYKEVYTDMKVEPEAAAYAP
                     PPPAKKPRKSTAEKPKVKEIIDERTRERLVYEVRQKCRNIEDICISCGSLNVTLEHPL
                     FVGGMCQNCKNCFLECAYQIDDDGYQSYCTICCGGREVLMCGNNNCCRCFVCEVDLL
                     VGPGAAQAAIKEDPWNCYMCGHKGTYGLLRRREDWPSRLQMFFANNHDQEFDPPKVYP
                     PVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYIASEVCEDSITVGMVRHQGKIM
                     YVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARKGLYCGTGRLFFEYYRLLHD
                     ARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMIDAKEVSAAHRARYFWGNL
                     PGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSNSIKQSKDQHFPVFMNEKE
                     DILWCTEMERVFGFPVHYTDVSNMSRLARQRLLGRSWSVPVIRHLFAPLKEYFACV"
ORIGIN

BASE COUNT   674 a   846 c   978 g   507 t

Query Match   70.0%; Score 3005; DB 9; Length 3005;
Best Local Similarity   100.0%;
Matches 3005; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
                Pred. No. 0;

Qy   1 GCCGCGGCGCACCCAGGGCGCGGACCCGCCGCCGCGGCGGCGGCCGGCCGACCCCAGGGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCCGCGGCGCACCCAGGGCGCGGACCCGCCGCCGCGGCGGCGGCCGGCCGACCCCAGGGCC 60

Qy   61 ATCGAAGCCCCCACCCACCCAGGCTGACAGAGAGGCACCGTTCACCAGAGGCCTCAACACC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATCGAAGCCCCCACCCACCCAGGCTGACAGAGAGGCACCGTTCACCAGAGGGCTCAACACC 120

Qy   121 ATCTATGTTTAAGTTTTAACTCTCGCCTCCAAAGACCACAGATAATTCCTTCCCCAAAGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATCTATGTTTAAGTTTTAACTCTCGCCTCCAAAGACCACAGATAATTCCTTCCCCAAAGCC 180

Qy   181 CAGCAGCCCCCCACCCAGCCCCGCGCGCAGCCCCAGCCCTGCCTCCCCGGGCCCCCAGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGCAGCCCCCCACCCAGCCCCGCGCGCAGCCCCAGCCCTGCCTCCCCGGGCCCCCAGCAT 240

Qy   241 GCCCTTCCAGGGACCCCCCGGGGGACACCAGCAGCTCTTGCTGGGGAGGGGGAGGGACCGAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCCCTCCAGCGCGCCCCCGGGGACACCAGCAGCTCTTGCTGGGGAGGGGGAGGGACCGAAA 300

Qy   301 GGACGGAGAGGAGCAGGAGGAGCCCGGCGTTGGCAAGGAGGAGAGCGCCAAGAGCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGACGGAGAGGAGCAGGAGGAGCCCGGCGTTGGCAAGGAGGAGAGCGCCAAGAGCCACCAC 360

Qy   361 GGCACCGGAAGGTGGGGGCGGGGCCCTGGGAAGGGCAGCAAGCGCAGCACCCCCGGTGGCAGAAGCGGTGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGCACCGGAAGGTGGGGGCGGGGCCCTGGGAAGGGCAGCAAGCGCAGCACCCCCGGTGGCAGAAGCGGTGA 420

Qy   421 CACGCCAAAGGACCCTGCGGTGATCTCCAAGTCCCCATCCATGGCCCAGGACTCAGGGCGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CACGCCAAAGGACCCTGCGGTGATCTCCAAGTCCCCATCCATGGCCCAGGACTCAGGGGCC 480

Qy   481 CTCAGAGCTATTACCCAATGGGAGACTTGGAGAAGCGGAGTGAGCCCCCAGCCAGGAGGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CTCAGAGCTATTACCCAATGGGGACTTGGAGAAGCGGAGTGAGCCCCCAGCCAGGAGG 540

Qy   541 GAGCCCCTGCTTGGGGGACACAAGGGGCCCCAGCAGAGAGGGCGGGGGTGCACTGAGAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAGCCCTGCTGTGGGGGCAGAAGGGGCGGGGGCCCCAGCAGAGAGGGGTGCACCTGAGAC 600

Qy   601 CCTGCCTGAAGCCTCAAGACCAGTGAAAATCGGCTGCTGCACCCCAAGGAGGGGCCGAGG 660

THIS PAGE BLANK (USPTO)

```
Db   601 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   660
         CCTGCCTGAAGCCTCAAGAGACCAGTGGAAAATGGCTGCTGCACCCCCAAGGAGGGCCGAGG

Qy   661 AGCCCCTGCCAGAAGACGGGGCAAAAGAACAGAAGGAGAGACCAACAGTGGAAAAATGGA   720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AGCCCCTGCCAGAAGCGGGGCAAAAGAACAGAAGGAGAGACCAACATCGAATCCATGAAAATGGA   720

Qy   721 GGGCTCCCGGGGCCGCTCGCGGGGTGGCTTGGGCTCCAGCCTCAGCGGCC   780
         ||||||||||||||||||||||||||||||||||||||||
Db   721 GGGCTCCCGGGGCCGCTCGCGGGGTGGCTTGGGCTCCAGCCTGTCAGCGGCC   780

Qy   781 CATGCCGAGGCTCACCTTCCAGGCGGGGGGGGACCCCTACTACTACATCAGCAAGCCAAGCGGGA   840
         ||||+|||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CATGCCGAGGCTCACCTTCCAGGCGGGGGGGGACCCCTACTACATCAGCAAGCCAAGCGGGA   840

Qy   841 CGAGTGGCTGGCACGCCTGGAAAAGGGGCTGAGAGAAGGTCATTGCAGGAAT   900
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CGAGTGGCTGGCACGCCTGGAAAAGGGGCTGAGAGAAGGTCATTGCAGGAAT   900

Qy   901 GAATGCTGTGGAAGAAAACCAGGGGCCCGGGGAGTCTCAGAAAGGTGGAGGAGGCCAGCCC   960
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GAATGCTGTGGAAGAAAACCAGGGGCCCGGGGAGTCTCAGAAAGGTGGAGGAGGCCAGCCC   960

Qy   961 TCCTGCTGTGCAGCAGCCCACTGACCCCGCATCCCCACTGTGCTACCACGCCTGAGCC   1020
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TCCTGCTGTGCAGCAGCCCACTGACCCCGCATCCCCACTGTGCTACCACGCCTGAGCC   1020

Qy  1021 CGTGGGGGTCCGATGCTGGGGGACAGAGAATGCCACCAAAGCAGGGGCCGGATGAGAGCCAGAGTA   1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CGAGGACGGCCAGCCGGGGCCCGGGGGCTTTGGCATTGGGGGAGTCTGGGGGGAAACTGCGGGGCTTCTC   1080

Qy  1081 CGAGGACGGCCAGCCGGGGCCCGGGGGCTTTGGCATTGGGGGAGTCTGGGGGGAAACTGCGGGGCTTCTC   1140
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1081 CTGGTGGCCAGGCCGCCAATGTGTCTTCTGGGTGGATGACGGGGTGTGGGAGCAGCTGAAGG   1140

Qy  1141 CTGGTGGCCAGGCCGCCAATGTGTCTTCTGGGTGGATGACGGGGTGTGGGAGCAGCTGAAGG   1200
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CTGGTGGCCAGGCCGCCAATGTGTCTTCTGGGTGGATGACGGGGTGTGGGAGCAGCTGAAGG   1200

Qy  1201 CACCCGCTGGGTGTCATGTGGTTCGGGAGACGGGCAAATTCTCAGTGGTGTGTGTTTTTGAGAAGCT   1260
         |||||||||||+|||||||||||||||||||||||||||||||||
Db  1201 CACCCGCTGGGTCATGTGGTTCGGGAGACGGGCAAATTCTCAGTGGTGTGTGTTTTTGAGAAGCT  1260

Qy  1261 GATGCGCCGCTGAGCTCGTTTTTGCAGTGGCGTTCCACCAGGCCACTGTACAACAAGCAGCCCAT   1320
         ||||+|||||||||||||||||||||||||||||||||||||||||||
Db  1261 GATGCCCGCTGAGCTCGTTTTTGCAGTGGCGTTCCACCAGGCCACTGTACAACAAGCAGCCCAT   1320

Qy  1321 GTACCGCAAAAGCCATCTACGAGAGTTCCTGCGACGTCCAGGTGGCCGCAGCACCGCGGGGGAAGCTGTTT   1380
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1321 CCCGGGTGTTGCCACGACACGGATGAGAGTGACACTGCCAAGGCCGTGGAGGTGCAGGAACAA   1380

Qy  1381 CCCGGGTGTTGCCACGACACGGATGAGAGTGACACTGCCAAGGCCGTGGAGGTGCAGGAACAA   1440
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CCCGGGTGTTGCCACGACACGGATGAGAGTGACACTGCCAAGGCCGTGGAGGTGCAGGAACAA   1440

Qy  1441 GCCCATGATTGAATGGGGCCCTGGGGGGGCTTCCAGCCTTCTCTGGGCCTCTGGAGCC   1500
         |||||||||||||||||||||||||||||||||||||||||
Db  1441 GCCCATGATTGAATGGGGCCCTGGGGGGGCTTCCAGCCTTCTCTGGGCCTCTGGAGCC   1500

Qy  1501 ACCAGAGAAGAGAAGAAATCCCTACAAAGAAGTGTACACGGACAATGTGGGTGGGAACCTGA   1560
         |||||||||||||||||||||||||||||||||||||||||||
Db  1501 ACCAGAGAAGAGAAGAAATCCCTACAAAGAAGTGTACACGGACAATGTGGGTGGGAACCTGA   1560

Qy  1561 GGCAGCTGCCTACCGCACCTCCACCAAAAAAGCCCCGGAAAGAGACACAGCGGGAGAA   1620
         |||+|||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GGCAGCTGCCTACCGCACCTCCACCAAAAAAGCCCCGGAAAGAGACACAGCGGGAGAA   1620

Qy  1621 GCCCAAGGTCAAGGAGATTATTGATGAGCGCCACACCAAGAGAGAGCGGCGGCTGGTTGTACGAGGTGCG   1680
         ||||||||||||||||||||||||||||||||||||||||||
Db  1621 GCCCAAGGTCAAGGAGATTATTGATGAGCGCCACACCAAGAGAGAGCGGCGGCTGGTTGTACGAGGTGCG   1680

Qy  1681 GCAGAAGTGCCGGAACATTGCATCTCTGTGTGGGAGCCTCAAATGTTACCCT   1740
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1681 GCAGAAGTGCCGGAACATTGAGGACATCGGGAACATCTGCATCTCCTGTGTGGGAGCCTCAAATGTTACCCT   1740
```

```
Db  1681 GCAGAAGTGCCGGAACATTGAGGACATCGGGAACATCTGCATCTCCTGTGTGGGAGCCTCAAATGTTACCCT   1740

Qy  1741 GGAACACCCCCTCTTCGTTGGGAGGAATGTGCCAAAACTGCAAGAACTGCTTTTCTGGAGTG   1800
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 GGAACACCCCCTCTTCGTTGGGAGGAATGTGCCAAAACTGCAAGAACTGCTTTTCTGGAGTG   1800

Qy  1801 TGCGTACCAGTACGAGACGACGACGGCTACCAGTCCTACTGCACCATCTGTCTGTGGGGGCCG   1860
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TGCGTACCAGTACGAGACGACGACGGCTACCAGTCCTACTGCACCATCTGTCTGTGGGGGCCG   1860

Qy  1861 CATGTGCGGCACAAAGGGTACCTAGGGCTCGTGCCGCGCGAGGAGACTGGGCCCTCCCG   1920
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1861 CATGTGCGGCACAAAGGGTACCTAGGGCTCGTGCCGCGCGAGGAGACTGGGCCCTCCCG   1920

Qy  1921 TGGAGGTGCTCATGTGCGGAAAACAACAACTGCCTGCAGGTGCTTTTTGCGGTGGGA   1980
         ||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CCTCTTGGTTGGGGGGCCGGGGGGGCTGCCCAGGGAGCAGCCAATTAAAGGAAGACCCCTGGAACTGCTA   1980

Qy  1981 CATGTGGGGCACAAAGGGTACCTACGGGCGTGCTGCGGCCGAGGAGGAGACTGGGCCCTCCCG   2040
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1981 CATGTGGGGCACAAAGGGTACCTACGGGCGTGCTGCGGCCGAGGAGGAGACTGGGCCCTCCCG   2040

Qy  2041 GCTCCAGATGTTCTTCGCTAATAAACCACGACCAGGAATTTGACCCTCAAAAGGTTTTACCC   2100
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 GCTCCAGATGTTCTTCGCTAATAAACCACGACCAGGAATTTGACCCTCAAAAGGTTTTACCC   2100

Qy  2101 ACCTGTCCCAGCTGTGAGAAGAGGAAGCCCATCTCGGGTGCTGCGTGTCTCTTCTTTTGATGGAATCGC   2160
         |||||||||||||||||||||||||||||||||||||||||
Db  2101 ACCTGTCCCAGCTGTGAGAAGAGGAAGCCCATCTCGGGTGCTGCGTGTCTCTTCTTTTGATGGAATCGC   2160

Qy  2161 TACAGGGGTCCTCGGTGCTGAAGGACTTGGGCATTCAGGTGGCCTACATTGCCTCGGA   2220
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2161 TACAGGGGTCCTCGGTGCTGAAGGACTTGGGCATTCAGGTGGCCTACATTGCCTCGGA   2220

Qy  2221 GGTGTGTGAGGAGGACTCCATCACGGTGCCGCCACCAGGGGAGAGATCATGTTACGT   2280
         ||||||||||||||||||||||||||||||||||||||||||
Db  2221 GGTGTGTGAGGAGGACTCCATCACGGTGCCGCCACCAGGGGAGAGATCATGTTACGT   2280

Qy  2281 CGGGGAAGCGTCCGAGCGTCACACAGAAGCATATCCAGGGGGGGCCCCATTCGAATCTGGT   2340
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2281 CGGGGACGCGTCCCGCAGCGTCACACAGAAGGATATCCAGGAGTGGGGGCCCCATTCGGATCTGGT   2340

Qy  2341 GATTGGGGGCCAGTCCCTGCAAATGACCTCTCCATCGTCAACCCTGCTCCCGGAAGGGCCTCTA   2400
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2341 GATTGGGGGCAGTCCCTGCAAATGACCTCTCCATCGTCAACCCTGCTCCCGGAAGGGCCTCTA   2400

Qy  2401 CGAGGGCACTGGCCGGCTCTTCTTTGAGTGTTGTACCGGCCTCGCATGATGCGGGCCCAA   2460
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2401 CGAGGGCACTGGCCGGCTCTTCTTTGAGTGTTGTACCGGCCTCGCATGATGCGGGCCCAA   2460

Qy  2461 GGAGGGAGAGATGATCGCCCCTTCTTCTGGCCTCTTTTGAGAATGTGTGGGCCGGTTAG   2520
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GGAGGGAGAGATGATCGCCCCTTCTTCTGGCCTCTTTTGAGAATGTGTGGGCCGGTTAG   2520

Qy  2521 TGACAAGAGGGACACATCTCCGGATTTCTCGAGTCCAACCCCTGTGATTGATTGCCAAAGA   2580
         |||||||||||||||||||||||||||||||||||||||||||
Db  2521 TGACAAGAGGGACACATCTCCGGATTTCTCGAGTCCAACCCCTGTGATTGATTGCCAAAGA   2580

Qy  2581 AGTGTCAGCTGCACACAGGGCCGCTACTTCTGGGGTAACCTTCCCGGTATGAACAGGCC   2640
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2581 AGTGTCAGCTGCACACAGGGCCGCTACTTCTGGGGTAACCTTCCCGGTATGAACAGGCC   2640

Qy  2641 GTTGGCATCCACTGTGAATGATAAGCTGCAGGAGCTGCAGGAT   2700
         ||||||||||||||||||||||||||||||||||||||||||
Db  2641 GTTGGCATCCACTGTGAATGATAAGCTGCAGGAGCTGCAGGAT   2700

Qy  2701 AGCCAAGTTCAGCAAAGTGAGGACCATTACTACGAGGTCAAAACTCCCATAAAAGAGGGCAA   2760
         |||||||||||||||||||||||||||||||||||||||||||||
Db  2701 AGCCAAGTTCAGCAAAGTGAGGACCATTACTACGAGGTCAAAACTCCCATAAAAGAGGGCAA   2760

Qy  2761 AGACCAGCCATTTCCTGTGTCTTCATGAATGAGAAAGAGGAGACACATCTTATGGTGCACTGAAAT   2820
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2761 AGACCAGCCATTTCCTGTGTCTTCATGAATGAGAAAGAGGAGACACATCTTATGGTGCACTGAAAT   2820
```

THIS PAGE BLANK (USPTO)

Tue Jul 22 16:21:56 2003   us-09-720-0

```
Qy    2821 GGAAAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTCTCCAACATGAGCCGGCTTGGC 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 GGAAAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTCTCCAACATGAGCCGGCTTGGC 2880

Qy    2881 GAGGCAGAGACTGCTGGGCCCGGTCATTGGAGCGTGCCAGTCATCCGCCACCTCTTCGCTCC 2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 GAGGCAGAGACTGCTGGGCCCGGTCATTGGAGCGTGCCAGTCATCCGCCACCTCTTCGCTCC 2940

Qy    2941 GCTGAAGGAGTATTTTGCGTGTGTGTAAGGGACATGGGGGCAAACTGAGGTAGCGGACACA 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 GCTGAAGGAGTATTTTGCGTGTGTGTAAGGGACATGGGGGCAAACTGAGGTAGCGGACACA 3000

Qy    3001 AAGTT 3005
           |||||
Db    3001 AAGTT 3005
```

RESULT 3
AF068675

THIS PAGE BLANK (USPTO)